This Page Is Inserted by IFW Operations and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents will not correct images, please do not report the images to the Image Problem Mailbox.



LOW 07/344258

Results file low344-fig1-pir. res made by maryh on Wed 17 Apr 91 11:19:26-PDT.

Query sequence being compared: Number of sequences searched: Number of scores above cutoff: LOW344-FIG1. PEP 17731

3826

Results of the initial comparison of LOW344-FIG1. PEP with: Data bank : PIR 25.0, all entries

100000-

N -

U50000-

M

B -

R.

0 ->

F10000-

s -

E 5000-

G - >

U -

E

N -

Ē *

S 1000-

- *

500-

_

_

-

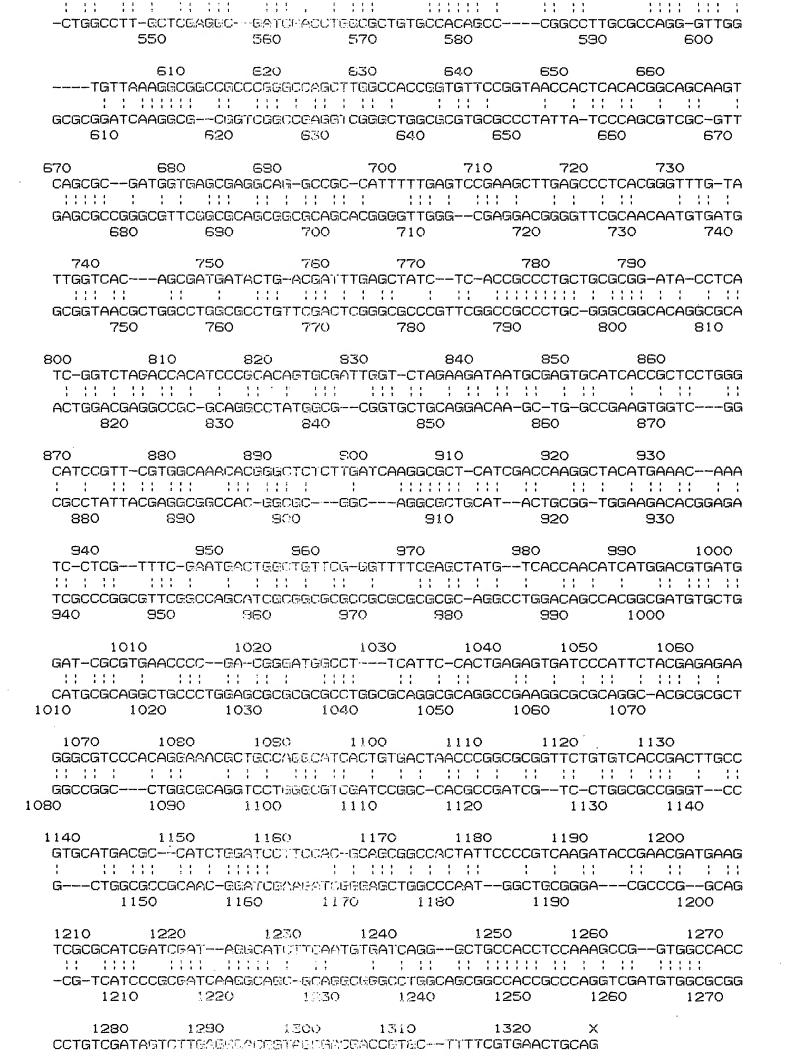
100-

_

50- *

_

10-



equipme	*									
5-										
_										
-										
-0004										
									×	
0										
1:	:::::		!	•	!		•	!	:	
SCORE O:	::25:	50	76	10 i	126	151	177	202	227	
			10		1720	101	111	202		
STDEV -1	9									

F'ARAMETERS

Similarity matrix Mismatch penalty Gap penalty Gap size penalty Cutoff score Randomization group	Unitary 1 1.00 0.05 6 0	K-tuple Joining penalty Window size	2 20 32
Initial scores to save Optimized scores to sav	20	Alignments to save	10
	/e 20	Display context	10

SEARCH STATISTICS

Scores:	Mean မိ	Median 7	Standard Deviation 2.65
Times:	CFU :03:03.05		Total Elapsed 00:09:14.00
Number of residues: Number of sequences sea Number of scores above (4595063 17731 3826	
Cut-off raised to 7. Cut-off raised to 8. Cut-off raised to 9.			

The scores below are sorted by initial score. Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Seque	nce Name	Description	Length	Init. Score	•	Sig.	Frame
		**** 83 standard deviatio			: *		
1.	A28214	Phosphotriesterase - Pseudo **** 4 standard deviatio			251 **	83. 28	. 0
2.	VGBEGX	Secreted glycoprotein gX - 1	Pse 498	19	47	4. 90	0
. 3.	SYBSYX	TyrosinetRNA ligase - Bac	111 419	19	43	4. 90	0
4.	NOHUG	Enclase gamma - Human #EC-n	umb 434	18	62	4. 52	. 0
5.	A27124	Hamsporting ATPase - Le	ish 974	18	68	4. 52	. 0
6.	S02077	Enolase gamma - Human (frag	men 433	18	62	4: 52	. 0
7.	A24742	Enolase gamma chain - Rat #	EC- 434	18	63	4. 52	. 0
8.	A24405	Ice nucleation protein - Ps	eud 1200	. 17	63	4. 15	0
9.	QDBP4L	Hypothetical protein D-206	- B 206	17	36	4. 15	0
10.	A28852	Histone H3(1) - Tetrahymena	py 135	17	27	4. 15	0
11.	B24255	Charlon of 88 A protein L12	pr 132	17	28	4. 15	0

		4444 3 standerd deviations ab	ove mean	***			
12.	S01921	Hypothetical protein 1 - Chlam	451	16	19	3. 77	0
13.	S01022	Hypothetical prutein P-2 - Chl	86	16	19	3.77	0
14.	HSXL32	Histone H3.2 - African clawed	135	16	30	3. 77	0
15.	S03605	Surface glycoprotein CD14 prec	366	16	61	3. 77	0
16.	A24225	Transducin beta chain - Bovine	340	16	40	3. 77	0
17.	A24853	Transducin beta chain, liver -	340	16	40	3. 77	0
18.	A26066	Segmentation protein eve - Fru	376	16	35	3.77	0
19.	A25457	Transducin beta chain — Bovine	340	16	40	3.77	0
20.	DAHUAL	Arachidonate 5-lipoxygenase -	674	16	65	3. 77	0

The scores below are sorted by optimized score. Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

				Init.	Opt.		
Seque	nce Name	Description	Length	Score	Score	Sig. 1	Frame
		**** 74 standard deviations	above me	ean ***	×		
1.	A28214	Phosphotriesterase - Pseudomomome *** 5 standard deviations		227 ean ***	251	74. 70	0
2.	WMBEBH	72K protein - Bovine herpesvin	r 664	1 1	72	5. 42	0
3.	QQBEE3	HHLF1 protein - Cytomegaloviro		9	70	4.64	0
4.	@@BEC3	HORF1 protein - Cytomegalovir			69	4. 26	0
5.	A25902	65K protein antigen - Mycobac		15	69	4. 26	0
6.	S05506	Phosphoenolpyruvate carboxyla	966	8	69	4. 26	0
7.	S00893	Adenylate cyclase precursor -	1706	10	69	4. 26	0
8.	502389	Cyclolysin - Bordetella pertus	s 1706	10	69	4. 26	0
		**** 3 standard deviations	above me	ean ***	×		
9.	A27124	H+-transporting ATPase - Leisl	າ 974	18	68	3. 87	0
10.	WFHUM	Mullerian inhibiting factor p	r 560	13	68	3. 87	0
11.	GNNY5P	Genome polyprotein - Poliovir	J 2207	13	68	3.87	0
12.	S04255	Regulatory protein qa-15 - Ne	u 918	8	68	3. 87	0
13.	GNWVWV	Genome polyprotein - West Nile	9 3430	10	68	3. 87	0
14.	QQBE8	Hypothetical BPLF1 protein - B	3149	10	67	3. 48	0
15.	B28894	Myeloperoxidase H7 - Human	830	11	67	3. 48	0
16.	OKBOG	Protein kinase, cEMP-dependen	t 67 0	11	67	3. 48	0
17.	DCZYPC	Pyruvate decarboxylase - Zymon	n 559	8	67	3. 48	0
18.	S02386	cyaB protein - Bordetella per	t 712	10	67	3. 48	0
19.	VGBEPB	Glycoprotein gIII precursor -	479	9	67	3. 48	0
20.	S00896	Ferredoxin—nitrite reductase	594	8	67	3. 48	0

1. LOW344-FIG1. PEP

A28214	Phosphotriesterase - Pseudomonas diminuta MG plasm
ENTRY	A28214 %Type Protein
TITLE	Phosphotriesterase - Pseudomonas diminuta MG plasmid pCMS1
SOURCE	Pseudomunas diminuta
ACCESSION	A20214
REFERENCE	(Sequence translated from the DNA sequence)
#Authors	McDaniel C. S., Harper L. L., Wild J. R.
#Journal	J. Bacterio). (1989) 170:2306-2311
#Title	Cloning and sequencing of a plasmid-borne gene (opd)
	encoding a phosphotriesterase.
SUPERFAMILY	#Name pluosphotriesterase
KEYWORDS	esturase
COMMENT	THIS SEIGUENCE HAS NOT BEEN COMPARED TO THE
	NUCLEOTIDE TEXNISLATION.
SUMMARY	#Mclecular-wellsh. 35563 #Length 325 #Checksum 539

```
227 Optimized Score =
Initial Score
                                             251
                                                 Significance = 74.70
Residue Identity =
                                                 Mismatches
                    76%
                       Matches
                                             260
                                                                  51
                     28 Conservative Substitutions
                                                                   0
                    20
                             30
                                                                   70
  - X
                                      40
                                                50
                                                         60
   MQTRRVVLKSAAAGTLLGGLAGCATWLDRSAQAIGSIRARPITISEAGFTLTHEDICGSSAGFLRAWPEFFG
   M@TRRVVLKSAAAPTLLGGLAGCATWLDRSA@AMRSIRARPITISEAGFTLTHEDI---SAAR@DSCVLG@S
                    20
                             30
                                      40
                                                50
   X
           10
                                                            60
                                   100
                                                     120
         80
                   90
                                            110
   SRKALAEKAVRGLR-ARAAGVR----TIVD-----VSTFDIGRDVSLLAEVSRAADVHIVAATGLWFDPPLSM
               ! ! ! ! ! ! ! ! ! ! !
                                       1 1 1
                                             ; ;
                                                          SSVAØSSSGKGCERIARØSGWRANDCRCVDFRYRSRRØFIGR----GFAGCRR----SYLAATGLWFDPPLSM
                                                       120
                     90
                             100
                                      110
     140
                   150
                            160
                                      170
                                               180
                                                        190
                                                                  200
   RLRYVEELT-----@FFLREI@YGIEDTGIRAGIIKVATTGKATPF@ELVLKAAARASLATGVPVTTHTAAS
                       ;
                            RLRYVEELTLVLPAVRFNMASKY----TGIRAGIIKVATTGKATPFQELVLKAAARASLATGVPVTTHTAAS
      140
               150
                            150
                                      170
                                               180
                                                        190
                                                                  200
                 220
                          230
                                    240
                                             250
                                                      260
        210
   @RDGERGRPPFLSPKLEPSRVCIGHSDDTDDLSYLTALLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSW@
   @RDGERGRPPFLSPKLEPSRVCIGHSDDTDDLSYLTALLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSW@
        210
                 220
                          230
                                    240
                                             250
                                                      260
               290
      280
                        300
                                  310
                                           320
                                                  X 330
   TRALLIKALID@GYMK@ILVSNDWLFGFSSYVTNIMDVMDRVNPDGMAFIPLRVIPFYERR
   TRALLIKALID@GYMK@ILVSNDWLFGFSSYVTNIMDVMDRVNPDGMAFIH
      280
               290
                        300
                                  310
                                           320
2. LOW344-FIG1, PEP
  WMBEBH
             72K protein - Bovine herpesvirus (type 2, strain
ENTRY
                         #Type Protein
TITLE
               72K protein - Bovine herpesvirus (type 2, strain
                 BMV)
               31-Mar-1990 #Sequence 31-Mar-1990 #Text 31-Mar-1990
DATE
PLACEMENT
                        1.0
                               2.0
                                     1.0
                                            1.0
SOURCE
               bovine mamillitis virus, bovine herpesvirus 2
               B29242
ACCESSION
REFERENCE
               (Sequence translated from the DNA sequence)
               Hammerschmidt W., Conraths F., Mankertz J., Pauli
   #Authors
                 G. , Ludwig H. , Buhk H. J.
   #Journal
               Virology (1988) 165:388-405
               Conservation of a gene cluster including
   #Title
                 glycoprotein B in bovine herpesvirus type 2
                 (BHV-2) and herpes simplex virus type 1 (HSV-1).
   #Comment
               The amino acid sequence is not given in this paper.
COMMENT
               The DNA sequence was obtained from GenBank, release
                 61.0.
COMMENT
               This virus is a member of the family Herpesviridae.
               #Name herpesvirus infected cell protein ICP18.5
SUPERFAMILY
             #Molecular-weight 72367 #Length 664 #Checksum 8190
SUMMARY
SEQUENCE
Initial Score
                                              72
                                                                5.42
               =
                     11
                        Optimized Score
                                       ===
                                                  Significance =
                    26%
                       Matchies
                                                                  203
Residue Identity =
                                             101
                                                  Mismatches
```

Conservative Substitutions

0

83

Gaps

```
1.0
                            20
                                      30
                                               40
           M@TFRVVLKSAAAGTLL-----GGL--AGCATWLDRSA@AIGSIRARPITISEA-GFTLTHEDI
            RLAGKICDHVT@@ARVRLDADEMF33NLFHVV@LSEARRARALHALEVSSKMTEANSGGPAEAPGPAAAQE---
          X 240
                     2350
                                      270
                              260
             70
    60
                     80
                              30
                                     100
                                               110
                                                          120
   CGSSAGFLRAWPEFFGSRKALAEKAVRGLR-ARAAGVRTIVDVSTFDIGRD-VSLLAEVSR----AADVHIV
     REASA-LLDAHHVFKSAPFGL--YAVSELRFWLSSGDRT--SGSTVDAFADNLSALAERERRYETGAVAVEL
                               330
                        320
                                          340
                   140
                          150
                                    160
                                             170
   AATG---LWFDPPLSMR---LRYVEELT@FFLREI@YGIEDTGIRAGIIKVATTGKATPF@ELVL----
   AAFGRRGEHFDRTFGDRVASI_DMVPAL----FVGG9SAAPDD@IEALVRACYNHHLSAP----VLRQLAGSE
                320
                           390
                                    400
                                            410
         190
                 200
                             210
                                      220
                                               230
                                                       240
   --KAAARASLATGVPVTTHTA-ASØRDG ----ERGRPPFLSPKLEPSRVCIGHSDDTDDLSYLTALLRGYLI
     HGDAEALRSALEG----LHAAEDPPGDGNAEKEARRAPSL-----GGGPEDDWAALAARAAADVGARR
                440
                        450
                                460
                                                470
                   270
                                         290
            260
                            280
                                                   300
                                                           310
   GLDHIPHSAIGLEDNASASPLLGIRSW@TRALLIKALID@GY----MK@ILVSN--DWLFGFSSYVTNIMD
   RL----YADRLYKRSLAS---LGROVREGRGELEKMLRVSTYGEVLPTVFAAVCNGFAARTRFCELTARAGT
                                       530
                                                    540
                   500
                           510
                                    520
                 330
                          340
   VMD-RVNPDGMAFIPLRVIPFYERRAS-HPKRCGASL
   VIDNRGNPD--TFDTHR----FMRASLMRHRVDPALLPGITHGFFE
 550
           560
                      570 580
3. LOW344-FIG1. PEP
  00BEE3
            HHLF1 protein - Cytomegalovirus (strain AD169)
ENTRY
              00BEE3 #Type Protein
TITLE
             HHLF1 protein - Cytomegalovirus (strain AD169)
DATE
              30-Sep-1963 #Sequence 30-Sep-1989 #Text 31-Dec-1989
              1358.0 1.0 1.0 2.0
PLACEMENT
SOURCE
              human cytomegalovirus, human herpesvirus 5
ACCESSION
              (Sequence translated from the DNA sequence) ·
REFERENCE
              Weston K., Barrell B.G.
   #Authors
              J. Mol. Biol. (1926) 192:177-208
   #Journal
   #Title
              Sequence of the chart unique region, short repeats,
               and part of the long repeats of human
               cytomedalitvirus.
COMMENT
              The DNA sequence was obtained from EMBL, release 13.
COMMENT
              This virus is a member of the family Herpesviridae.
GENETIC
                HHLF1
   #Name
              *Name cytomegalovirus HORF1 protein
SUPERFAMILY
            #Molecular-weight 83981 #Length 788 #Checksum 7858
SUMMARY
SEQUENCE
                  9 Optimized Score =
Initial Score
           --
                                         70 Significance =
                                                         4.64
                 23% Motches =
Residue Identity =
                                         90
                                            Mismatches
                                                          241
Gaps
                  55 Conservative Substitutions
                                                            0
                                  30
                  10
                         20
                                                40
                                                         50
           M@TPRVVLKS/(AAS) (LLEGLAGCATVLDRSA-@AIG----SIRARPITISEAGFTLTHED
```

```
ASAPHPASLLTAVRRHLNGRUCCGMLALGAVLPARVLGCAAGPATGTAAGTTSPPAASGTETEAAGGDAPCA
               340
                      350 360
                                         370
                                                 380
               70
                           FIO
                                    90
                                            100
       60
                                                    110
                                                            120
   ICG--SSAGFLRAWPEFFGSRKA----LA-ENAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAEVSRAADVH
   IAGAVGSAVPVPP@PYGAAGGGAICVPNADAHAVVGADAAAAAAPTVMVGSTAMAGPAAS--GTVPRAMLVV
           410 420 430
                                    440
          130
                 140
                         150
                                   160
                                            170
   IVAATGLWF-DPPLSMRLRYVEELT@FFLREI@YGIEDTG-IRAGIIKVATTGKATPF@E-----LVLKA
       LLDELGAVFGYCFLDGHVYPLAAELSHFLRAGVLGALALGRESAPAAEAARRLLPELDREGWERPRWDALHL
              480 490
                           500
                                       510
                                                520
                               220
      190
              200
                       210
                                       230
                                                240
                                                            250
   AARASLATGVPVTTHTAASQRDGERGRPPFLSPKLEPSRVCIGHSDDTDDLSYLTALLR---GYLIGLD--HI
     HPRAALWAREP-HG@LAFLL NPG-RGEAEVLTLATKHPAICANVEDYL@D-----ARRRADA@ALGLDLATV
              550
                       560
                               570
                                       580
                        270
                                280
                                        290
                                                 300
               260
                                                         310
   PHSAIG-----LEDNASASPILGIRSW@TRALLIKALID@GYMK@ILVSNDWLFGFSSYVTNIMDVMD
                   VMEAGG@MIHKKTKKPKGKEDESLMKGKHSRYTR-PTEPPLTP@ASLGRALRRDDEDWKPS-----RLPGED
         610 620 630
                                 640 650
                                                660
                330
                        340
   RVNPDGMAFIPL---RVIPFYERR-ASHRKRCQASL
         SWYDLDETFWVLGSNRKNDVY@RRWKKTVLRCGLEIDRPMPTVPKG
            680
                 690
                         700 X
4. LOW344-FIG1. PEP
            HQRF1 protein - Cytomegalovirus (strain AD169)
  QQBEC3
ENTRY
                       #Type Protein
TITLE
              HORF1 protein - Cytomegalovirus (strain AD169)
              30-Sep-1989 #Sequence 30-Sep-1989 #Text 30-Sep-1989
DATE
PLACEMENT
                     1.0 1.0 1.0
                                       1.0
SOURCE
              human cytomegalovirus, human herpesvirus 5
ACCESSION
              C26078
              (Sequence translated from the DNA sequence)
REFERENCE
              Weston K. , Barrell B. G.
   #Authors
   #Journal
              J. Mol. Biol. (1986) 192:177-208
              Sequence of the short unique region, short repeats,
   #Title
               and part of the long repeats of human
               cytomegalovirus.
COMMENT
              The DNA sequence was obtained from EMBL, release 13.
COMMENT
              This virus is a member of the family Herpesviridae.
GENETIC
                H@RF1
   #Name
SUPERFAMILY
              #Name cytomegalovirus HQRF1 protein
            #Molecular-weight 91047 #Length 846 #Checksum 2604
SUMMARY
SEQUENCE
Initial Score
                                            Significance =
                   9 Optimized Score =
                                         69
                                                          4.26
Residue Identity =
                  22% Matches
                                         86
                                            Mismatches
                                                           248
                                                            0
Gaps
                  46 Conservative Substitutions
                          20
                                  30
                                                40
                                                         50
           M@TRRVVLKSAAAGTLLGGLAGCATWLDRSA-@AIG----SIRARPITISEAGFTLTHED
```

ASAPHPASLLTAVRRHLN@RLCCGWLALGAVLPARWLGCAAGPATGTAAGTTSPPAASGTETEAAGGDAPCA

330	X 340	350	360	370	380	390
60	70	80	90	100	110	120
	LRAWPEFFGSRKA	ALA-EKA		VRTIVDVSTF		
: ; ;; IAGAVGSAVP 400	VPPBPYGAAGGGA 410	AICVPNADAHA		APTVMVGSTA	MAGPAAS——G 450	TVPRAMLVV 460
	0 140 DPPLSMRLRYVEE					180 I VI KA
: :	: :	; ; ;	:	: :	1	;
LLDELGAVEG 470	YCPLDGHVYPLAA 480				LPELDREWWE 520	RPRWDALHL 530
	200 PVTTHTAASQRDG					
1 1 1	1 1	1 1 1	;	4 † † 1	1	:: :
	PHGQWEFMFF 550					IGFYØIRKP 600
	270 SPLLGIRSW@T					0
:	1 :	: :	:	: ;	:	: :
	——QPPPPSR@TKF 610 620		GABASVSYALR 640			
	330 34 -VIPFYERRASHF					
: :	1 1	4 1				
DSRPDDIKGR 670 68	RLLKATØRRGAE] O 690		CYDBRFT			
5. LOW344-FIG1 A25902	.PEP 65K protein	antigen -	Mycobacter	ium leprae	÷	
ENTRY	A25302	SType Pro	otein			
TITLE	65K prute:	in antigen	- Mycobact	erium lepr	ae`	
SOURCE ACCESSION	A25902	rium lapred	:			
REFERENCE #Authors		translated Sweetser D			nce)	
#Journal	Proc. Nat.	Acad. Sci	. USA (198	6) 83:7013		_
#Title SUMMARY	#Molecular-				leterminant ecksum 304	
SEQUENCE					,	
Initial Score Residue Identi	ty = 23%	Dutimized Matches Conservati	ت ت	90 Mis	nificance matches	= 231
Gaps						= 0
X MGT	10 RRVVLKSAA——3	20 NATELINGLAGO	30 ATWLDRSAGA		O ARPITISE	50 AGFTLTHED
!	: : ASCGPFSRALHPA	1 1	;	:: :	:: :	; ; ;
10		30	V 1	40	50	60 60
60	70			00	110	120
! !	AWPEFFGSRKALE 	1 1 1 1	1 1	1 1 1	1 1	1 1
GLNSLADAVK 70	VTLGPKG-FMVVL 80	TITAKUW WB <mark>.</mark> SO	NDGVSIAKEI 100	ELEDPYEKIG	iAELVKEV- 120	-AKKTDDVA
130 ATGLWFDPPL	140 SMRLRYVEELITGE					190 KAAARASLA
:	1 1		1 1 1	1 1 1	1	11 1

```
140
                         1 36.
                                  160
                                            170
                                                      180
          200
                                       350
                            210
                                                   230
                                                            240
   1 1 1
                              1 1 1 1
   AGDQSIGDLIAEAMDKYGNEGVITVEESN FGLQLEL FEGMRFDKGYISGYFVTDAERQEAVLEEPYILLVS
                       550
                                230
                                          240
          260
                    270
                             230
                                       250
                                                 300
                                                          310
                                                                      320
   IPHSAIGLEDNASASPLLGÜRSVOTPALLIKALIDQGYMKQILVSNDWLFGFSSYVTNIMDVMDR---VNPDG
                             1 1 1 1
                                                    !!
                  : ! !
                                     .
                                            :: :
   SKVS----TVKDLLFLLEKVI@AGKGLLIIAGDVEGEALSTLVVNKIRGTFKSVAVKAPGFGDRRKAML@D
 270
                280
                          290
                                   300
                                             310
                                                       320
                                                                330
          330
                    340
                           \mathcal{A}'
   MAFIPLRVIPFYERRASHRKRCGASL
               •
   MA-ILTGAQVISEEVQLTLENTDLSLLCKARKVVMT
     340
               350
                        360
                                  370
  LOW344-FIG1. PEP
  S05506
              Phosphoenolpyruwate carboxylase 1 - Common ice pla
ENTRY
                S05508
                          *Type Protein
TITLE
                Phosphoenolpyruvete carboxylase 1 - Common ice plant
                  #EC-number 4.1.1.31
SOURCE
                Mesembryantherum crystallinum #Common-name common
                  ice plant
ACCESSION
                S05508
REFERENCE
                Cushman J. C. . Bohnsht H. J.
   #Authors
                Nucleic Acids Res. (1989) 17:6745
   #Journal
                Nucleotide sequence of the gene encoding a CAM
   #Title
                  specific isoform of phosphoenolpyruvate
                  carboxylase from Mesembryanthemum crystallinum.
   #Reference-number S05506
                S05506
   #Accession
   #Molecule-type DNA
   #Residues
               1-986 (CUS)
REFERENCE
   #Authors
                Rickers J., Cushman J. C., Michalowski C. B., Schmitt
                  J. M., Bohnert H. J.
   #Journal
                Mol. Gen. Genet. (1989) 215:447-454
   #Title
                Expression of the CAM-form of phospho(enol)pyruvate
                  carboxylase and nucleotide sequence of a full
                  length cDNA from Mesembryanthemum crystallinum.
   #Reference-number 502716
   #Accession
                502716
   #Molecule-type mRNA
                1-856 (RIC)
   #Residues
 SUMMARY
             #Molecular-weight 110659 #Length 966 #Checksum
                                                              725
SEQUENCE.
Initial Score
                      8 Optimized Score
                                                69
                                                   Significance =
                                                                   4. 26
                ŧ-..
                                         ===
                                                94
Residue Identity =
                     23%
                        Matches
                                          122
                                                    Mismatches
                                                                    212
Gaps
                     90
                         Conservative Substitutions
                                                                      0
                     10
                              30
                                        30
                                                        40
                                                                 50
             M@TPRVVLKSAAAGT!LLORLASCATVLDRSAQA----IGSIR----ARPITISEAGFTLTHEDI
                                      : :
                                          ! !
                                                   1:1
   SVRRSLL@KHGRIEDCLAGLYARDITPDDL@ELDEAL@REI@AAFRTDEIRRT@PTP@DEMRAGMSYFHETI
    180
              190
                       200
                                 210
                                           220
                                                     230
                                                              240
```

90

100

110

70

60

```
COSSAGELRAWPEFFGSRKALAEKHVRGLRAR------AAGVRTIVDVSTFDIGRDVSLLAEVSR
       WNGVPKFLR------RLDTALK-NIGITERVFYNAPLI@FSSWMGGDRDGNPRVTPEVTRDVCLLA-RMM
               260 270 280 290
  250
           130
                       140
                            150
                                      160
  AADVHIVAATGLWFDPPLSM-------PLR-YVEELT@FFLREI@YGIEDTGIRAGIIKVATTGKATPF@E--L
   ; ; ;
   AANMYFSQIDELMF--ELSMWRCTDELRERASELHKYSKRDSKHYIE-----FWKQIPSSEPYR
                 330
                      340
                                350
                           200 210
 180
                 190
                                          220
                                                   230
        VILADVRDKLYYTRERSR@LLASEVSEIPVEATFTEID@-----FLEPLELCYRSLCACGDRPVADGS
                                               420
          380 390
                        400
                                      410
           250
                              270
                                                     300
                       250
                                     280
                                             290
  YL----TALLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSW@TRALLIKALID@GYMK@ILVSNDWLFGF
   LLDFMR@VATFGLCLVKLDIR@ESERHTDVMDAITTHLGIGS--YRDWTEEKR@D--WLLSELRGKRPLFGP
        440 450 460 470
                                       480
       310
              320
                      330
                              340
  SSYVT-NIMDVMDRVNPDGMAFIPLEV1FFY-ERRASHRKRCQASL
      DLPRTDEIADVLDTIN--VIAELPEDSFGAYVISMATAPSDVLAVELLQRECKVKK
         510 520 530 540 X
  500
7. LOW344-FIG1, PEP
  S00893
           Adenylate cyclase precursor - Bordetella pertussis
ENTRY
            S00893
                   - #Typs Protein
            Ademylate cyclase precursor - Bordetella pertussis
TITLE
              #EC-number 4, 6, 1, 1
INCLUDES
            probable haemolysin
SOURCE
            Bordetella pertussis
ACCESSION
            500893
REFERENCE
            Glaser P., Ledant D., Sezer O., Pichot F., Ullmann
  #Authors
              A. , Danchin A.
  #Journal
            Mol. Microbiol. (1983) 2:19-30
  #Title
            The calmodulin-sensitive adenylate cyclase of
              Bordetella pertussis: cloning and expression in
              Escherichia coli.
   #Reference-number S00893
  #Accession S00893
  #Molecule-type DNA
  #Residues 1-1706 (GLA)
  #Cross-reference EMBL: Y00545
GENETIC
  #Name
              cya
        #Molecular-weight 177506 #Length 1706 #Checksum 6461
SUMMARY
SEQUENCE
                10 Optimized Score =
                                     69 Significance = 4.26
Initial Score =
Residue Identity =
                                     93 Mismatches =
                23% Matches
                                                     206
Gaps
                97 Conservative Substitutions
                                                     , 0
                       20
               10
                               30
                                                40
          MØTRRVVLKSAAAGTLLIGGLAGCATWLDRSAØAIGSIR----ARPITI---SEA
           LMTQFGRAGSTNTP@EAASLSAAVFGL-GEASSAVAETVSGFFRGSSRWAGGFGVAGGAMALGGGIAAAVGA
                                              540
        490
               500
                   510 520
                                      530
                                                      550
```

```
60 70
                                80
                                        90
    50
                                               100
                                                       110
   GFTLTHEDICG--SSAGFLRAWPEFFG-----SRKALAEKAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAE
   GMSLTDDAPAG@KAAAGAEIAL@LT@GTVELASSIALA-------LAAARGVTSGL@VAGASAG------
       560 570 530
                                    590 600
            130 140 150
       120
                                                 160
   VSRAADVHIVAATGLWFDPPL/SMRLRYVEELTQFFLREIQYGIE-----DTGIRAGIIKVATT
     ---AAAGALAAALSPMEIYGLV@@SHYAD@LDKLA@ESSAYGYEGDALLA@LYRDKTAAEGAVAGVSAVLST
    610 620 630 640 650 660
        180
               190
                       200
                                   210
   GKATPF@ELVLKAAARASLATGVPVTTHTAAS@RDG-----ERGRPPFLSPKLEPSRVCIGHSDDTDDLSYL
   VGA-----AVSIAAA-AS-VVGAPVAVVT--SLLTGALNGILRGV@@PIIEKL----ANDYARKIDELGGP
   680
                690
                   700 710 720
        240
               250
                               270
                                      280
                                                       300
                       260
                                               290
   TA-----LLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSWØTRALLIKALIDØGYMKØILVSNDWLFGF
        @AYFEKNL@ARHE@LANSDGLRKMLADL@AGWNASSVIG----V@TTEISKSAL----ELAAITGNADNL--K
    740 750 760 770
                                     780
                                               790
              3:20
                        330
       310
                                340
   SSYVTNIMDVMDRVNPDGMAFIP--LEVIPFYERRASHRKRC@ASL
   SVDV----FVDRFV@GERVAG@PVVLDVAAGGIDIAS-RKGERPALTFITPLAAPG
             810 820 830 840
  800
8. LOW344-FIG1. PEP
  S02389 Cyclolysin - Bordetella pertussis
ENTRY
             S02389 #Type Protein
TITLE
             Cyclolysin - Bordetella pertussis
INCLUDES
             adenylate cyclase #EC-number 4.6.1.1\ hemolysin
SOURCE
             Bordetella pertussis
             502389
ACCESSION
REFERENCE
   #Authors
             Glaser P., Sakamoto H., Bellalou J., Ullmann A.,
              Danchin A.
   #Journal
             EMBO J. (1988) 7:3997-4004
   #Title
             Secretion of cyclolysin, the calmodulin-sensitive
              adenylate cyclase-haemolysin bifunctional protein
              of Bordetella pertussis.
   #Reference-number 502386
   #Accession S02389
   #Molecule-type DNA
   #Residues 1-1706 (GLA)
GENETIC
   #Name
              CYAA
             THIS SEQUENCE HAS NOT BEEN COMPARED TO THE
COMMENT
              NUCLEOTIDE TRANSLATION.
FEATURE
   1-312
                      #Domain adenylate cyclase (ADE)\
   313-1706
                     #Domain hemolysin (HEM)
SUMMARY
         #Molecular-weight 177476 #Length 1706 #Checksum 6271
SEQUENCE
Initial Score =
                10 Uptimized Scure = 69 Significance = 4.26
Residue Identity =
Gaps =
                 23% Motches
                                 ===
                                       93 Mismatches
                                                       206
                 97 Cariservative Substitutions
                                                         0
Gaps
                 10
                        50
                                30
                                                  40
          MQTPRVVLKSAAAGTLLGGLAGCATWLDRSAQAIGSIR-----ARPITI---SEA
```

```
LMT@FGRAGSTNTP。EBAASLEAAVEEL、GEASBAVAETVSGFFRGSSRWAGGFGVAGGAMALGGGIAAAVGA
                500 510
                              520
                                        530
             60
                    70
                                 80
                                        90
                                                100
    50
                                                        110
   GFTLTHEDICG---SSAGFLRAWPERFG-----SRKALAEKAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAE
   GMSLTDDAPAG@KAAAGAEIAL@LTGGTVELASSIALA-----LAAARGVTSGL@VAGASAG-----
                                    590 600
           570 590
                      140
              130
                              150
                                                  160
   VSRAADVHIVAATGLWFDPPLSMRLRYVEEL1@FFLREI@YGIE-----DTGIRAGIIKVATT
     ---AAAGALAAALSPMEIYGI.V@@5HYAD@LDKLA@ESSAYGYEGDALLA@LYRDKTAAEGAVAGVSAVLST
         620 630 640 650 660 G
                                    210 220
        180 190
                        200
   GKATPF@ELVLKAAARASLATGVPVTTHTAAS@RDG-----ERGRPPFLSPKLEPSRVCIGHSDDTDDLSYL
    VGA----AVSIAAA-AS-VVGAPVAVVT--SLLTGALNGILRGVQQPIIEKL----ANDYARKIDELGGP
                        700 710 720
                250 260 270 280
                                            290
        240
                                                         300
   TA-----LLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSWØTRALLIKALIDØGYMKØILVSNDWLFGF
         @AYFEKNL@ARHE@LANSDGLRKMLADL@AGWNASSVIG---V@TTEISKSAL----ELAAITGNADNL--K
    740 750 760 770
                                      780
                                                 790
              320
                       330 340
   SSYVTNIMDVMDRVNPDGMAFIP---LRVIPFYERRASHRKROGASL
   SVDV----FVDRFV@GERVAG@PVVLDVAAGGIDIAS-RKGERPALTFITPLAAPG
             810
                 320 830
                                  840
9. LOW344-FIG1. PEP
  A27124
        H+-transporting ATPase - Leishmania donovani
                     #Type Protein
 ENTRY
             H - transporting ATPase - Leishmania donovani
 TITLE
               #EC-number 3.6.1.35
 ALTERNATE-NAME proton-transporting ATPase
 SOURCE
             Leishmania Jonovani
 ACCESSION
             A27124
             (Sequence translated from the DNA sequence)
 REFERENCE
             Meade J. C., Shaw J., Lemaster S., Gallagher G.,
   #Authors
               Stringer J. R.
             Mol. Cell. Diol. (1987) 7:3937-3946
   #Journal
   #Title
             Structure and expression of a tandem gene pair in
               Leishmania domovani that encodes a protein
               structurally homologous to eucaryotic
               cation-transporting ATPases.
             The authors translated the codon AGA for residue 352
   #Comment
               as Lys.
 SUMMARY
           #Molecular-weight 107476 #Length 974 #Checksum 834
 SEQUENCE
Initial Score = 13 Optimized Score = 68 Significance = Residue Identity = 24% Matches = 93 Mismatches = Gaps = 82 Conservative Substitutions =
                                                        3.87
                                                         210
Gaps
                             20
                                     30
           MØTRR----VVLKSAAAGTLLGGLAGCATWLDRSAØAIGSIRARPITISEAGFTLTHE---
                FLDPPRPDTKDTTRRSKEYGYDVKHTT60HLLTAKEMO-RMLDLDPNILTADKLP@IKDANDLPEDLGEKYG
          510 520 530 540 550
  500
```

60 DICGSSAGFLRA	WPEFFGSRKA	ALAEKAVRGLA	ARAAGV		GRDVSLLAEVS	RAADVHIVA
; ; ;; DMMLSVGGFA@V 570 580	FFE	HKEMIVETLE	GRIGYTICAMT		RADVGIAVH	
	140 RLRYVE ;; VVVEAML\	15 	O KEIBYGIEDT ; ; ; SFLTYRISAT	160 GIRAGIIK ¦	170 VATTGKA ; ;; FSLTPKAYGSV	TPFQE :: 'DPHFQFFHL
180 190 LVLKAAARASLA ;; ; PVLMFMLITLLN 710	TGVPVT1 	THTAASORDGE	:RGREPFLSP ;;;; :KP@KWNL-P	KLEPSRVCIGH ; VVFVS	ISDDTDDLSYLT ; ;	ALLRGYLIG :
LLWIGLE	DNASASPLLO ; GYSSQY\	RIRSWOTR 	ALLIKALI ; ; _AQLPQGKLV	D@GYMK@ILVS ; ; ; ;	NDWLFGFSSY\ 	/TNIMDVMDR 「GGHFFFYMP
VNPDGMAFIP ; ;; PSPILFCGAIIS 820 830	LRVIPFYERF ; ; LLV———STMA		BASL ; DNVLTEGLAV			
10. LOW344—FIG1. WFHUM		inhibiting	, factor p	recursor -	Human	
ALTERNATE—NAME DATE PLACEMENT SOURCE ACCESSION REFERENCE #Authors	Mullerian Mullerian 13-Aug-19 596. 0 Homo sapi A01397 (Sequence Cate R. L. Farber Gash D. Torres R. C., R	n inhibitir n inhibitir 186 #Sequer 5.0 l. lens #Commo e translate , Mattalia N.M., Cheu J., Chow E G., Wallns	ig factoring substantice 13-Aug O 1.0 on-name mand and from the and R. J., ung A., Ni E.P., Fish or B.P., R	-1986 #Text 1.0	ence) Tizard R., Tey A.Z., ertonis J.M.	9
#Journal COMMENT	This prot factor and B c	beta, inhi	ologous t bin alpha area of	o transform chain, and best homolo proteins.	inhibin A	
COMMENT	Although binding	it does no g sites, Ml	ot compete S cen inh	with EGF f ibit the		
GENETIC	•			EGF recept	OF IN VICE	J.
#Introns SUPERFAMILY KEYWORDS	#Name in∤ testicula antitum	er glycopro	steirN gon	2 adal differ duct\ TGF-		og/
FEATURE 1-25		*Ocmain	signal an	d propeptio	de sequence	
		(STG):	_		•	

(STG)%

26-560	Marotein Mullerian inhibiting factor
64,329	(MAT)\ #Binding-site carbohydrate (Asn)
SUMMARY #Molecular-we	(potential) Bight 59192 #Length 560 #Checksum 3812
SEQUENCE	,
Initial Score = 13 (Optimized Score = 68 Significance = 3.87 Matches = 99 Mismatches = 211
	Conservative Substitutions = 0
X 10	
	AGTLLG-GLAGCATWL-DRSAQA-IGSIRARPIT-ISEAGFTL
LPGAQSLCPSRDTRYLVLAVDRPA 190 200	AGAVRGSGLALTLØPRGEDSRLSTARLØALLFGDDHRCFTRMTPALLLL 210 220 230 240 250
	80 90 100 110
	RKALAEKAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAEVSR-AA
PRSEPAPLPAHGQLDTVPFPPPRF	PSAELEESPPSADPFLETLTRLVRALRVPPARASAPRLAL
260 270	280 290 300 310
	140 150 160 170 180 _RYVEELT@FFLREI@YGIEDTGIRAGIIKVATTGKATPF@ELVLK
	: :: :: :: :: :: : : : : : : : : : : :
320 330 34	
	210 220 230 240
	RDGERGRPPFLSPKLEPSRVCIGHSDDTDDLSYLTALLRGYLIGLD-
AAAAELRSLPGLPPATAPLLA 390 400	RLIALCP-GGPGGLGDPLRALLLKALQGLRVEWRGRDP 410 420 430 440
250 260 270	280 290 300 310
	RSW@TRALLIKALID@GYMK@ILVSNDWLFGFSSYVTNIMDV
	_SVDLRAERSVLIPETY@ANNC@GVCGWP@SDRNPRYG
430 460	470 400 400 500
7.00	7.40
MDRVNPDGMAFIPLRV	30 X IPFYER-RASHRKRCØASL ; ; ; ; ;

X 560

```
Query sequence being compared:
                                     LOW344-FIG1, PEP
Number of sequences searched:
                                               15409
Number of scores above cutoff:
                                                3758
      Results of the initial comparison of LOW344-FIG1. PEP with:
   Data bank : Swiss-Prot 14, all entries
100000-
Ν
U50000-
Μ
В
Ε
R
0
F10000-
S
E 5000-
Ø
U
Ε
N
C
Ε
S 1000-
      *
   500-
```

100-

Results file low344-fig1-spt res made by maryh on Wed 17 Apr 91 11:34:12-PDT.

_										
_										
10-	×									
5-										
_										
_										
_										
-										
_									×	
0									······································	
::	1 1 1 1 1	;	:	:	;) !	!	:	;	
SCORE 0:	1:25:	50	76	101	126	151	177	202	227	
STDEV -1	9									

PARAMETERS

Similarity matrix Mismatch penalty Gap penalty Gap size penalty Cutoff score Randomization group	Unitary 1 1.00 0.05 6	K-tuple Joining penalty Window size	2 20 32
Initial scores to save Optimized scores to sav	30	Alignments to save Display context	10 10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	6	7	2.69
Times:	CPU 00:03:15,00		Total Elapsed 00:13:53.00

Number of residues: 4914263 Number of sequences searched: 15409 Number of scores above cutoff: 3758

Cut-off raised to 7. Cut-off raised to 8. Cut-off raised to 9.

The scores below are sorted by initial score. Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Seque	nce Name	Description		Init. O Score S	•	Sig. Fi	rame
1.	OPD\$PSEDI	**** 82 standard deviation PHOSPHOTRIESTERASE (EC 3.5. **** 4 standard deviation	325	227	251	82. 28	0
2.	MTC3\$CHVN1	MODIFICATION METHYLASE CVI	9 II 377	19	25	4.84	0
3.	SYY\$BACCA	TYROSYL-TRNA SYNTHETASE (EC	C 6. 419	19	43	4.84	0
4. 5.	VGLX\$PRV ENOG\$RAT	SECRETED GLYCUPROTEIN GX (C GAMMA ENGLASE (EC 4.2.1.11)		19 18	47 63	4. 84 4. 47	0

6.	ENOG\$HUMAN	GAMMA FINULASE (EC 4.2.1.11) (2	433	18	62	4, 47	O
7.	ATXA\$LEIDO	PROBABLE E1-E2 TYPE CATION ATP	974	18	68	4.47	Q
8.	I CEN\$ERWHE	ICE NUCLEATION PROTEIN (GENE N	1258	18	63	4. 47	0
9.	ATXB\$LEIDO	PROBABLE E1-E2 TYPE CATION ATP	974	18	68	4.47	0
10.	H31\$TETPY	HISTONE H3.1.	135	17	27	4.10	0
11.	KCCA\$RAT	CALCIUM/CALMODULIN-DEPENDENT P	478	17	57	4.10	0
12.	KCCA\$MOUSE	CALCIUM/CALMODULIN-DEPENDENT P	478	17	57	4.10	Q
13.	ICEN\$PSESY	ICE NUCLEATION PROTEIN (GENE N	1200	17	63	4.10	0
14.	CHA2\$BOMMO	CHORION CLASS A PROTEIN L12 PR	132	17	28	4.10	0
15.	Y206\$LAMBD	HYPOTHETICAL PROTEIN ORF206.	206	17	36	4.10	0
		**** 3 standard deviations ab	ove mear	7 ****			
16.	ARSA\$HUMAN	ARYLSULFATASE A PRECURSOR (EC	507	16	64	3.72	0
17.	HMEV\$DROME	SEGMENTATION PROTEIN EVEN—SKIP	376	16	35	3. 72	0
18.	GBB1\$HUMAN	GUANINE NUCLEOTIDE-BINDING PRO	340	16	40	3.72	0
19.	H32\$XENLA	HISTONE H3, 2,	135	16	30	3.72	0
20.	CD14\$MOUSE	CD14 DIFFERENTIATION ANTIGEN P	366	16	61	3.72	0

The scores below are sorted by optimized score. Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

				Init.	Opt.		
Seque	nce Name	Description	Length	Score	Score	Sig. F	=rame
		**** 76 standard deviations	above me	ean ***	*		
1.	OPD\$PSEDI	PHOSPHOTRIESTERASE (EC 3.5		227	251	76, 20	0
		**** 4 standard deviations		an ***	X		
2.	YHL1\$HCMVA	HYPOTHETICAL PROTEIN HHLF1.	788	9	70	4. 74	0
3.	Y@R1\$HCMVA	HYPOTHETICAL PROTEIN HORF1.	846	9	63	4. 34	0
4.	CYAA\$BORPE	CALMODULIN-SENSITIVE ADENYLATE	1706	10	69	4. 34	0
5.	CAP1\$MESCR	PHOSPHOENOLPYRUVATE CARBOXYLAS	966	8	69	4. 34	0
		**** 3 standard deviations	above me	an ***	*		
6.	MIS\$HUMAN	MULLERIAN INHIBITING FACTOR PR	560	13	68	3. 95	0
7.	ATXA\$LEIDO	PROBABLE E1-E2 TYPE CATION ATF	974	18	68	3.95	0
8.	POLG\$WNV	GENOME POLYPROTEIN (CAPSID PRO	3430	10	68	3.95	0
9.	ATXB\$LEIDO	PROBABLE E1-E2 TYPE CATION ATF		18	68	3. 95	0
10.	ATPO\$0ENBI	ATP SYNTHASE ALPHA CHAIN, MITC	511	9	68	3. 95	0
11.	POLG\$POL2L	GENOME POLYPROTEIN (COAT PROTE	2207	13	68	3 . 95	0
12.	PGKH\$WHEAT	PHOSPHOGLYCERATE KINASE, CHLOR	480	8	. 68	3. 95	0
13.	EXON\$HSV11	ALKALINE EXONUCLEASE (EC 3.1.1	626	9	67	3. 55	0
14.	KGP\$BOVIN	CGMP-DEPENDENT PROTEIN KINASE	670	11	67	3. <i>5</i> 5	0
15.	KGPB\$HUMAN	CGMP-DEPENDENT PROTEIN KINASE,	686	11	67	3, 55	0
16.	VGL3\$PRV	GLYCOPROTEIN GIII PRECURSOR.	479	9	67	3, 55	0
17.	PHYB\$ARATH	PHYTOCHROME B (GENE NAME: PHYE	1172	9	67	3.55	0
18.	ATI1\$HSV11	ALPHA TRANS-INDUCING FACTOR 73		9	67	3, 55	Ö
19.	PGCA\$RAT	CARTILAGE-SPECIFIC PROTEOGLYCA		9	67	3. 55	ō
20.	PYR1\$YEAST	CARBAMOYL-PHOSPHATE SYNTHETASE		1 1	67	3. 55	Õ

1. LOW344-FIG1. PEP

OPD\$PSEDI PHOSPHOTRIESTERASE (EC 3.5, -. -) (GENE NAME: OPD).

```
ID
     OPD$PSEDI
                    STANDARD;
                                    PRT;
                                           325 AA.
AC
     P13739;
     01-JAN-1990 (REL. 13, CREATED)
DT
DT
     01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT
     01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
DE
     PHOSPHOTRIESTERASE (EC 3.5. -. -) (GENE NAME: OPD).
OS.
     PSEUDOMONAS DIMINUTA.
OG
     PLASMID PCMS1.
```

- OC PROKARYOTA; BACTERIA; GRAM-NEGATIVE AEROBIC RODS AND COCCI;
- OC PSEUDOMONADACEAE.

```
[1] (STRAIN MG, SECLIENCE FRUM N. A.)
RN
RA
     MCDANIEL C.S., HARPER L.L., WILD J.R.;
     J. BACTERIOL. 170:2306-2311(1988).
RL
CC
     -!- PATHWAY: PESTICIDE DETOXIFICATION.
     -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC
     EMBL; M20392; PPPTE.
DR
DR
     PIR; A28214; A28214.
     PLASMID; HYDROLASE; MEMBRANE.
KW
               325 AA;
                       35583 MW;
                                  528207 CN;
SØ
     SEQUENCE
Initial Score
                    227
                        Optimized Score
                                             251
                                                 Significance = 76.20
Residue Identity =
                    76%
                        Matches
                                             260
                                                 Mismatches
                                                                  51
                     28
                        Conservative Substitutions
                                                                   0
Gaps
                                                                  70
           10
                    20
                             30
                                      40
                                                50
                                                         60
   MQTRRVVLKSAAAGTLLGGLAGCATVLDRSAQAIGSIRARPITISEAGFTLTHEDICGSSAGFLRAWPEFFG
   MQTRRVVLKSAAARTLLGGLAGCATVLDRSAQAMRSIRARPITISEAGFTLTHEDI---SAARQDSCVLGQS
           10
                    20
                             30
                                      40
                                               50
                                                           60
   X
         80
                   90
                                   100
                                            110
                                                     120
                                                              130
   SRKALAEKAVRGLR-ARAAGVR----TIVD----VSTFDIGRDVSLLAEVSRAADVHIVAATGLWFDPPLSM
               : : :
                                            : :
                                                          SSVAØSSSGKGCERIARØSGWRANDCRCVDFRYRSRRØFIGR----GFAGCRR----SYLAATGLWFDPPLSM
                     30
                             100
                                      110
                                                      120
                                                               130
  70
           80
     140
                   150
                            160
                                     170
                                               180
                                                        190
                                                                 200
   RLRYVEELT-----@FFLREI@YGIEDTGIRAGIIKVATTGKATPF@ELVLKAAARASLATGVPVTTHTAAS
            RLRYVEELTLVLPAVRFNMASKY----TGIRAGIIKVATTGKATPF@ELVLKAAARASLATGVPVTTHTAAS
      140
               150
                            160
                                     170
                                              180
                                                        190
                                                                 200
        210
                 220
                          230
                                   240
                                             250
                                                      260
                                                               270
   QRDGERGRPPFLSPKLEPSRVCIGHSDDTDDLSYLTALLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSWQ
   @RDGERGRPPFLSPKLEPSRVCIGHSDDTDDLSYLTALLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSW@
        210
                 220
                          230
                                    240
                                             250
                                                      260
                                                               270
               290
                                           320
      280
                        300
                                  310
                                                 X 330
   TRALLIKALID@GYMK@ILVSNDWLFGFSSYVTNIMDVMDRVNPDGMAFIPLRVIPFYERR
   TRALLIKALID@SYMK@ILVSNDWLFGFSSYVTNIMDVMDRVNPDGMAFIH
      280
               290
                        300
                                  310
                                           320
  LOW344-FIG1. PEP
  YHL1$HCMVA HYPOTHETICAL PROTEIN HHLF1.
                                 PRT;
 ID
     YHL1$HCMVA
                   STANDARDE
                                       788 AA.
 AC
     P09695;
 DT
                 (REL. 10, CREATED)
     01-MAR-1989
                 (REL. 10, LAST SEQUENCE UPDATE)
 DT
     01-MAR-1989
                 (REL. 13, LAST ANNOTATION UPDATE)
 DT
     01-JAN-1990
 DE
     HYPOTHETICAL PROTEIN HHLF1.
 08
     HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
 OC
     VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE.
 RN
     [1] (SEQUENCE FROM N. A.)
 RA
     WESTON K., BARRELL B. G.;
 RL
     J. MOL. BIOL. 192:177-208(1986).
     -!- SIMILARITY: TO HWL!1, HHLF5, HHLF6, HHLF7, AND HQRF1.
 CC
 DR
     EMBL; X04630; HEHCMVU.
 DR
     PIR; C27349; GGBEE3.
 KW
     HYPOTHETICAL PROTEIN.
     CARBOHYD
 FT
                 763
                       76
                                PUTENTIAL.
 FT
     CARBOHYD
                110
                       118
                                FOTENTIAL.
```

PUTENTIAL.

FT

CARBOHYD

223

```
788 AA; 85961 MW; 2964529 CN;
SØ
    SEQUENCE
Initial Score =
                9 Optimized Score = 70 Significance = 4.74
Residue Identity =
                                    30 Mismatches =
                23% Matches
                            1.73
                                                     241
            E.E.
                55 Conservative Substitutions
                                                      0
                       20
                10
                               30
                                           40
          M@TRRVVLKSAAAGTLLGGLAGCATWLDRSA-@AIG-----SIRARPITISEAGFTLTHED
            ASAPHPASLLTAVRRHLNGRI.CCGWLALGAVLPARWLGCAAGPATGTAAGTTSPPAASGTETEAAGGDAPCA
       330 X 340 350 360 370 380
             70
                        80
                               90
                                      100 110
  , ICG--SSAGFLRAWPEFFGSRKA----LA-EKAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAEVSRAADVH
   IAGAVGSAVPVPP@PYGAAGGGAICVPNADAHAVVGADAAAAAAPTVMVGSTAMAGPAAS--GTVPRAMLVV
          410 420 430 440 450
         130
               140
                     150
                               160
                                       170
   IVAATGLWF-DPPLSMRLRYVEELT@FFLREI@YGIEDTG-IRAGIIKVATTGKATPF@E-----LVLKA
      LLDELGAVFGYCPLDGHVYPLAAEI.SHFLRAGVLGALALGRESAPAAEAARRLLPELDRE@WERPRWDALHL
         480 490 500 510 520
                    210 220 230 240
            200
      190
   AARASLATGVPVTTHTAAS@RDGERGRPPFLSPKLEPSRVCIGHSDDTDDLSYLTALLR--GYLIGLD--HI
    HPRAALWAREP-HGQLAFLLRPG-RGEAEVLTLATKHPAICANVEDYLQD-----ARRRADAQALGLDLATV
       550 560 570 580
                                               590
    540
                                                      600
              260 270 280 290
                                            300
                                                   310
   PHSAIG-----LEDNASASPLLGIRSW@TRALLIKALID@GYMK@ILVSNDWLFGFSSYVTNIMDVMD
    VMEAGGQMIHKKTKKPKGKEDESLMKGKHSRYTR-PTEPPLTPQASLGRALRRDDEDWKPS-----RLPGED
        610 620 630 640 650 660
              330 340 X
   RVNPDGMAFIPL---RVIPFYERR-ASHRKRC@ASL
       SWYDLDETFWVLGSNRKNDVY@RRVKKTVLRCGLEIDRPMPTVPKG
        680 690 700 X 710
   670
3. LOW344-FIG1. PEP
  YOR1$HCMVA HYPOTHETICAL PROTEIN HORF1.
ID
    Y@R1$HCMVA STANDARD;
                          PRT; 846 AA,
AC
    P09715;
DT
    01-MAR-1989 (REL. 10, CREATED)
DT
    01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT
    01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
    HYPOTHETICAL PROTEIN HORF1.
DE
OS.
    HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
OC
    VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE.
RN
    [1] (SEQUENCE FROM N. A.)
    WESTON K. , BARRELL B. G. ;
RA
RL
    J. MOL. BIOL. 192:177-208(1986).
CC
    -!- SIMILARITY: TO HWLF1, HHLF1, HHLF5, HHLF6, AND HHLF7.
DR
    EMBL; X04650; HEHCMVU.
    PIR; C26070; G@BEC3.
DR
KW
    HYPOTHETICAL PROTEIN.
FT
    CARBOHYD
             76 76
                          POTENTIAL.
FT
            113
                  118
    CARBOHYD
                         POTENTIAL.
                  223
            223
    CARBOHYD
                         POTENTIAL.
 FT
    SEQUENCE 846 AA; 91047 MW; 3448605 CN;
SØ
```

```
Initial Score = 9 Optimized Score = 69 Significance =
                22% Matches
                            -----
Residue Identity =
                                      86 Mismatches =
Gaps
                48 Conservative Substitutions
                                                        0
                10
                        20
                                30
                                             40
                                                     50
          M@TRRVVLKSAAAGTLLGGLAGCATVLDRSA-@AIG-----SIRARPITISEAGFTLTHED
            ASAPHPASLLTAVRRHLNØRLCCGVLALGAVLPARVLGCAAGPATGTAAGTTSPPAASGTETEAAGGDAPCA
       330 X 340 350 360 370 380
                                 90
                         30
                                        100
             70
                                                110
   ICG--SSAGFLRAWPEFFGSRKA----LA-EKAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAEVSRAADVH
   IAGAVGSAVPVPP@PYGAAGEGAICVPNADAHAVVGADAAAAAPTVMVGSTAMAGPAAS--GTVPRAMLVV
         410 420 430 440 450
         130 140 150 160
                                         170
   IVAATGLWF-DPPLSMRLRYVEELTØFFLREIØYGIEDTG-IRAGIIKVATTGKATPFØE-----LVLKA
      LLDELGAVFGYCPLDGHVYPLAAELSHFLRAGVLGALALGRESAPAAEAARRLLPELDRE@WERPRWDALHL
      470 480 490 500 510 520 530
      190 200 210 220 230 240 250
   AARASLATGVPVTTHTAAS@RDGE-RGRFFFLSPKLEPSRVCIGHSDDTDDLSYLTALLRGYLIGLDHIPHS
    HPRAALWAREP---HG@WEFMFRE@RGDPINDPLAFRLSDARTLGLDLTTVMTER@S@LPEKYIGFY@IRKP
             550 560 570 580 590 600
                             290 300
           270
                      280
   AIGLEDNASASPLLGIRSW@T---RALLIKALID@GYMK@ILVSNDWLFGFSSYV----TNIMDVMDRVNP
     PWLME-----QPPPPSR@TKPDAATMPPPLSA@ASVSYALRYDDESWRPLSTVDDHKAWLDLDESHWVLG
                      630 640 650 660
          610 620
          330 340
   320
   D-GMAFIPLR-VIPFYERRASHRKRCGASL
   DSRPDDIK@RRLLKAT@RRGAEIDRPMPVVPEECYD@RFT
 670 680 690 700
4. LOW344-FIG1. PEP
  CYAA$BORPE CALMODULIN-SENSITIVE ADENYLATE CYCLASE PRECURSOR (
ID
    CYAA$BORPE STANDARD; PRT; 1706 AA.
AC
    P15318;
DT
    01-APR-1990 (REL. 14, CREATED)
DT
    01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
    01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
DT
DE
    CALMODULIN-SENSITIVE ADENYLATE CYCLASE PRECURSOR (EC 4.6.1.1)
DE
    (ATP PYROPHOSPHATE-LYASE) (ADENYLYL CYCLASE) (CYCLOLYSIN) (CONTAINS:
DE
    HEMOLYSIN) (GENE NAME: CYA).
08
    BORDETELLA PERTUSSIS.
OC
    PROKARYOTA; BACTERIA; GRAM-NEGATIVE AEROBIC RODS AND COCCI; UNCERTAIN.
RN
    [1] (STRAIN 18323, SEQUENCE FROM N.A.)
RA
    GLASER P., LADANT D., SEZER O., PICHOT F., ULLMANN A., DANCHIN A.;
    MOL. MICROBIOL: 2:19-30(1988).
RL
CC
    -!- FUNCTION: THIS ADENYLATE CYCLASE BELONGS TO A SPECIAL CLASS OF
CC
       BACTERIAL TOXIN, IT ACTS ON MAMMALIAN CELLS BY ELEVATING CAMP-
CC
       CONCENTRATION AND THUS DISRUPTS NORMAL CELL FUNCTION.
CC
    -!- CATALYTIC ACTIVITY: ATF = 3,'5'-CYCLIC AMP + PYROPHOSPHATE.
CC
    -!- SUBCELLULAR LOCATION: RELEASED EXTRACELLULARLY IN A PROCESSED
CC
       FORM.
CC
    -!- DISEASE: WHOOPING COUGH.
DR
    EMBL; Y00545; BPCYA.
DR
    PIR; S00893; S00893
```

```
PROSITE; PS00330; HEMDLYSIN_CALCIUM.
DR
KW
    LYASE; CAMP SYNTHESIS; HEMOLYSIS; TOXIN; VIRULENCE; WHOOPING COUGH;
KW
    CALCIUM-BINDING; REPEAT.
              1 1706
FT
                          CALMODULIN-SENSITIVE ADENYLATE
FT
                          CYCLASE PRECURSOR.
FT
    CHAIN
                          CALMODULIN-SENSITIVE ADENYLATE
FT
                          CYCLASE.
    CHAIN
          1 1300
                          HEMOLYSIN, BY SIMILARITY TO E. COLI
FT
                          HEMOLYSINS (HYLA).
FT
FT
    DOMAIN
             1
                 360
                          A, CALMODULIN-SENSITIVE CATALYTIC
FT
                          CENTRE.
           361
                 912
                          B, ALA/GLY RICH.
FT
    DOMAIN
            913 1656
                          C.
FT
    DOMAIN
    DOMAIN 1657 1706 D, ASP/GLY RICH.
FT
    SEQUENCE 1706 AA; 177505 MW; 1.25518E+07 CN;
SØ
Initial Score = 10 Optimized Score -
Residue Identity = 23% Matches =
97 Conservative Substitu
                10 Optimized Score = 69 Significance = 4.34
                                    93 Mismatches = 206
                97 Conservative Substitutions
                                                     0
                                               40
             10
                       20 30
          M@TRRVVLKSAAAGTLLGGLAGCATWLDRSA@AIGSIR-----ARPITI---SEA
          LMT@FGRAGSTNTP@EAASLSAAVFGL-GEASSAVAETVSGFFRGSSRWAGGFGVAGGAMALGGGIAAAVGA
        490 500 510 520 530 540 550
                              80 90 100
           60 70
   GFTLTHEDICG--SSAGFLRAWPEFFG----SRKALAEKAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAE
   GMSLTDDAPAG@KAAAGAEIAL@LTGGTVELASSIALA-----LAAARGVTSGL@VAGASAG-----
       560 570 580
                                  590 600
             130 140 150
                                              160
      120
                                                      170
   VSRAADVHIVAATGLWFDPPLSMRLRYVEELT@FFLREI@YGIE-----DTGIRAGIIKVATT
    ---AAAGALAAALSPMEIYGLV@QSHYADQLDKLAQESSAYGYEGDALLAQLYRDKTAAEGAVAGVSAVLST
    610 620 630 640 650 660 670
                                 210 220
        180 190 200
   GKATPF@ELVLKAAARASLATGVPVTTHTAAS@RDG----ERGRPPFLSPKLEPSRVCIGHSDDTDDLSYL
   VGA----AVSIAAA-AS-VVGAPVAVVT--SLLTGALNGILRGVØØPIIEKL----ANDYARKIDELGGP
              690 700 710 720
   680
        240 250 260 270 280 290 300
   TA-----LLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSW@TRALLIKALID@GYMK@ILVSNDWLFGF
   @AYFEKNL@ARHE@LANSDGLRKMLADL@AGWNASSVIG---V@TTEISKSAL----ELAAITGNADNL--K
    740 750 760 770
                                   780
          320
                      330 340 X
   SSYVTNIMDVMDRVNPDGMAFIP--LRVIPFYERRASHRKRCQASL
   SVDV----FVDRFV@GERVAG@PVVLDVAAGGIDIAS-RKGERPALTFITPLAAPG
            810 820 830 840
  800
5. LOW344-FIG1. PEP
  CAP1#MESCR PHOSPHOENOLPYRUVATE CARBOXYLASE 1 (EC 4.1.1.31) (G
    CAP1$MESCR STANDARD; PRT; 966 AA.
ID
AC
    P10490;
    01-JUL-1989 (REL. 11, CREATED)
DT
    01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
DT
DT
```

PHOSPHOENOLPYRUVATE CARBOXYLASE 1 (EC 4.1.1.31) (GENE NAME: PPCA).

DE

```
COMMON ICE PLANT (MESEMBRYANTHEMUM CRYSTALLINUM).
OC
    EUKARYOTA; PLANTA; SPERMATOPHYTA; ANGIOSPERMAE.
    [1] (SEQUENCE FROM N. A.)
RN
    RICKERS J., CUSHMAN J., MICHALOWSKI C., SCHMITT J., BOHNERT H. J.;
RA
    MOL. GEN. GENET. 215:447-454(1989).
RL
    [2] (SEQUENCE FROM N. A. )
RN
    CUSHMAN J. C. , BOHNERT H. J. ;
RA
    NUCLEIC ACIDS RES. 17:6745-6745(1989).
RL
    -!- FUNCTION: TO FORM OXALDACETATE, A FOUR-CARBON DICARBOXYLIC ACID
CC
       SOURCE FOR THE TRICARBOXYLIC ACID CYCLE.
CC
    -!- CATALYTIC ACTIVITY: ORTHOPHOSPHATE + OXALOACETATE = H(2)0 +
CC
CC
       PHOSPHOENOLPYRUVATE + CO(2).
    -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC
DR
    EMBL; X13660; MCPPCR.
    EMBL; X14587; MCPPCA.
DR
    LYASE; CARBON DIOXIDE FIXATION; ALLOSTERIC ENZYME;
KW
KW
    TRICARBOXYLIC ACID CYCLE.
    SEQUENCE 966 AA; 110659 MW; 4690045 CN;
SØ
Initial Score = 8 Optimized Score = 69 Significance = 4.34
Residue Identity = 23% Matches = 94 Mismatches = 212
Gaps = 90 Conservative Substitutions = 0
Gaps
                10
                        20
                                              40
                                30
           MØTRRVVLKSAAAGTLLGGLAGCATWLDRSAØA---IGSIR---ARPITISEAGFTLTHEDI
            SVRRSLL@KHGRIRDCLA@LYAKDITPDDK@ELDEAL@REI@AAFRTDEIRRT@PTP@DEMRAGMSYFHETI
         190 200 210 220 230 240
    180
                                      90 100
        70 80
    60
   CGSSAGFLRAWPEFFGSRKALAEKAVRGLRAR------AAGVRTIVDVSTFDIGRDVSLLAEVSR
   WNGVPKFLR-----RLDTALK-NIGITERVPYNAPLI@FSSWMGGDRDGNPRVTPEVTRDVCLLA-RMM
                260 270 280 290 300
  250
                       140 150 160 170
    120 130
   AADVHIVAATGLWFDPPLSM-----RLR-YVEELTGFFLREIGYGIEDTGIRAGIIKVATTGKATPFGE--L
   AANMYFSQIDELMF--ELSMWRCTDELRERAEELHKYSKRDSKHYIE-----FWKQIPSSEPYR
        320 330 340 350
                            200 210 220
 180
                190
            ----AARASLATGV----PV-TTHTAASQRDGERGRPPFLSP-KLEPSRVCIGHSDDTDDLS
   VILADVRDKLYYTRERSR@LLASEVSEIPVEATFTEID@-----FLEPLELCYRSLCACGDRPVADGS
   370 380 390 400
                                         410 420 430
        240 250 260 270 280 290 300
   YL----TALLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSWØTRALLIKALIDØGYMKØILVSNDWLFGF
   LLDFMRQVATFGLCLVKLDIRGESERHTDVMDAITTHLGIGS--YRDWTEEKRQD--WLLSELRGKRPLFGP
        440 450 460 470
                                         480 490
       310 320 330 340 X
   SSYVT-NIMDVMDRVNPDGMAFIPLRVIPFY-ERRASHRKRCQASL
      DLPRTDEIADVLDTIN--VIAELPSDSFGAYVISMATAPSDVLAVELL@RECKVKK
  500 510 520 530 540 X 550
6. LOW344-FIG1. PEP
  MISSHUMAN MULLERIAN INHIBITING FACTOR PRECURSOR (MIS).
   MIS$HUMAN STANDARD; PRT; 560 AA.
ID
```

OS.

AC

DT

P03971;

23-OCT-1985 (REL. 02, CREATED)

```
DT
    23-OCT-1986 (REL. 02, LAST SEQUENCE UPDATE)
DT
    01-AUG-1988 (REL. 08, LAST ANNOTATION UPDATE)
DE
    MULLERIAN INHIBITING FACTOR PRECURSOR (MIS).
    HUMAN (HOMO SAPIENS).
OS.
OC
    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC
    EUTHERIA; PRIMATES.
RN
     [1] (SEQUENCE FROM N. A. )
    CATE R. L., MATTALIANO R. J., HESSION C., TIZARD R., FARBER N. M.,
RA
    CHEUNG A., NINFA E.G., FREY A.Z., GASH D.J., CHOW E.P., FISHER R.A.,
RA
    BERTONIS J. M., TORRES G., WALLNER B. P., RAMACHANDRAN K. L.,
RA
    RAGIN R. C., MANGANARO T. F., MCLAUGHLIN D. T., DONAHOE P. K.;
RA
RL
    CELL 45:685-698(1986).
CC
     -!- FUNCTION: THIS GLYCOPROTEIN, PRODUCED BY THE SERTOLI CELLS OF THE
        TESTIS, CAUSES REGRESSION OF THE MULLERIAN DUCT. IT ALSO IS ABLE
CC
CC
        TO INHIBIT THE GROWTH OF TUMORS DERIVED FROM TISSUES OF MULLERIAN
CC
        DUCT ORIGIN.
CC
    -!- SUBUNIT: DIMER OF IDENTICAL CHAINS LINKED BY AN INTERCHAIN
CC
        DISULFIDE BOND.
CC
    -!- SIMILARITY: TO TGF-BETA, INHIBIN ALPHA AND BETA CHAINS.
CC
     -!- ALTHOUGH IT DOES NOT COMPETE WITH EGF FOR RECEPTOR BINDING SITES,
        MIS CAN INHIBIT THE AUTOPHOSPHORYLATION OF THE EGF RECEPTOR IN
CC
CC
        VITRO.
DR
    EMBL; KO3474; HSMIS.
DR
    PIR; A01397; WFHUM.
    PROSITE; PS00250; TGF BETA.
DR
KW
    GROWTH FACTOR; GLYCOPROTEIN; GONADAL DIFFERENTIATION; SIGNAL.
FT
    SIGNAL 1 18
                            PUTATIVE.
    PROPEP 19 25 PUTATIVE.
CHAIN 26 560 MULLERIAN
CARBOHYD 64 64 POTENTIAL.
CARBOHYD 329 329 POTENTIAL.
FT
                            MULLERIAN INHIBITING FACTOR.
FT
FT
FT
    SEQUENCE 560 AA; 59192 MW; 1428811 CN;
SØ
Initial Score =
                  13 Optimized Score = 68 Significance = 3.95
Residue Identity =
                  24% Matches
                                          99 Mismatches =
Gaps
                  92 Conservative Substitutions
                            20
                                           30
                                                    40
                   10
           M@TRRVVL-KSAAA@TLLG-GLA-----GCATVL-DRSAQA-IGSIRARPIT-ISEAGFTL
             LPGAQSLCPSRDTRYLVLAVDRPAGAVRGSGLALTLQPRGEDSRLSTARLQALLFGDDHRCFTRMTPALLLL
         190
                200 210 220 230
                             80 90 100 110
                  70
         60
   THEDICGSSA-GFLRAWPEFFGSRKALAE----KAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAEVSR-AA
          PRSEPAPLPAHG@LDTVPFPPPRPSAELEESPPSADPFLETLTRLVR-----ALRVPPARASAPRLAL
                                                 300
       260 270 280 290
                    140 150 160
              130
                                                170
   DVHIVA--ATGL--WFDPPLSMRLRYVEELT@FFLREI@YGIEDTGIRAGIIKVATTGKATPF@ELV---LK
   DPDALAGFP@GLVNLSDPAALERLLDGEEPLLLLLR---PTAATTGDPAPLHDPTSAFWATALARRVAAEL@
                                        360
             330
                      340 350
                200 210
         190
                                 220
                                          230
                                                   240
   AAA--RASLATGVPVTTHTAAS@RDGERGRPPFLSPKLEPSRVCIGHSDDTDDLSYLTAL--LRGYLIGLD-
   AAAAELRSLPGLPPATAPLLA-----RLLALCP-GGP----GGLGDPLRALLLLKAL@GLRVEWRGRDP
     390
              400
                            410
                                         420 430
                                           290
                   270
                              280
                                                  300
          260
   HIPHSA-IGLEDNASASP-LLGIRSWØTRA---LLI----KALIDØGYMKØILVSNDWLFGFSSYVTNIMDV
     RGPGRAGRSAGATAADGPCALRELSVDLRAERSVLIPETYBANNCBG-----VCGWPBSDRNPRYG
      450 460 470 480 490
```

```
320
                         330
                                             340
   MDRVNPDGM-----AFIPLRVIPFY----ER-RASHRKRC@ASL
      NHVVLLLKMBARGAALARPPCCVPTAYAGKLLISLSEERISAHHVPNMVATECGCR
                520 530 540
      510
                                             550
7. LOW344-FIG1. PEP
  ATXA$LEIDO PROBABLE E1-E2 TYPE CATION ATPASE 1A (EC 3.6.1.-).
     ATXA$LEIDO
 ID
                    STANDARD;
                                  PRT;
                                         974 AA.
 AC
     P11718;
DT
     01-OCT-1989 (REL. 12, CREATED)
     01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT
     01-OCT-1989 (REL. 12, LAST ANNOTATION UPDATE)
DT
     PROBABLE E1-E2 TYPE CATION ATPASE 1A (EC 3.6.1.-).
DE
os
     LEISHMANIA DONOVANI.
     EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; MASTIGOPHORA.
 OC
RN
     [1] (SEQUENCE FROM N. A.)
RA
     MEADE J.C., SHAW J., LEMASTER S., GALLAGHER G., STRINGER J.R.;
     MOL. CELL. BIOL. 7:3937-3946(1987).
RL
 CC
     -!- CATALYTIC ACTIVITY: ATP + H(2)0 = ADP + ORTHOPHOSPHATE.
 CC
     -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC
         (E1-E2 ATPASES).
     -!- SIMILARITY: THE TWO L. DONOVANI CATION-TRANSPORTING ATPASE GENES
 CC
 CC
         ARE 98% HOMOLOGOUS.
 CC
     -!- CAUTION: IN POSITION 351 THE N.A. SEQUENCE PREDICTS ARG, THE
 CC
         PROTEIN TRANSLATION SHOWN IN THREE PLACES IN THE PAPER GIVES
 CC
         LYS, WHICH IS CONSERVED IN ALL KNOWN E1-E2 ATPASES. WE HAVE
CC
         USED LYS AFTER CONFIRMATION FROM THE AUTHORS.
DR
     EMBL; M17889; LDCATP1.
DR
     PROSITE; PS00154; ATPASE E1 E2.
KW
     HYDROLASE; ATP HYDROLYSIS; TRANSMEMBRANE; PHOSPHORYLATION;
KW
     MAGNESIUM; ATP-BINDING.
                                 PUTATIVE.
FT
                 93
     TRANSMEM
                       112
     TRANSMEM 118
TRANSMEM 265
TRANSMEM 000
FT
                        137
                                 PUTATIVE.
FT
                       286
                                 PUTATIVE.
FT
                      321
     TRANSMEM 631 651
TRANSMEM 662 684
TRANSMEM 698 712
TRANSMEM 738 761
TRANSMEM 813 840
TRANSMEM 869 887
TRANSMEM 869 887
                                 PUTATIVE.
FT
                                 PUTATIVE.
FT
                                 PUTATIVE.
FT
                                 PUTATIVE.
FT
                                 PUTATIVE.
FT
                                 PUTATIVE.
 FT
                                 PUTATIVE.
FT
                                 PHOSPHORYLATION.
 SQ
     SEQUENCE
                974 AA; 107448 MW; 5115862 CN;
Initial Score
                     18 Optimized Score = 68 Significance =
                                                                   3.95
Residue Identity =
                     24%
                         Matches
                                          122
                                                93
                                                    Mismatches
                                                                    210
Gaps
                     82 Conservative Substitutions
                                                                      0
             X
                          10
                                   20
                                             30
                                                       40
             M@TRR----VVLKSAAAGTLLGGLAGCATWLDRSA@AIGSIRARPITISEAGFTLTHE---
                     FLDPPRPDTKDT1RRSKEYGVDVKM1TGDHLL1AKEMC-RMLDLDPN1LTADKLP@IKDANDLPEDLGEKYG
            510
                      520
                                          540
                                              550
  500
                          530
                                                              560
                                                         110
                 70
                          80
                                       90
                                               100
   DICGSSAGFLRAWPEFFGSRKALAEKAVRGLRAR---AAGVRTIVDVSTFDIGRDVSLLAEVSRAADVHIVA
       DMMLSVGGFAQVFPE-----HKFMIVETLR@RGYTCAMTGDGVNDAPALKRADV--GIAVHGATDAARAA
                                              610
           580
  570
                          590
                                   600
                                                         620
       130
               140
                                150
                                            160
                                                       170
```

ATGLWFDPPLSMRLRYVE-----ELT@FFLREI@YGIEDT---GIRAGIIKVATTGKA----TPF@---E

```
1 1
                     ADMVLTEPGLS----VVVEAMLVSREVF@RMLSFLTYRISATL@LVCFFFIACFSLTPKAYGSVDPHF@FFHL
                  650
                           660 670
                                             680
                                                       690
                                                                700
   180
          190
                       200
                               210 220
                                                 230
   LVLKAAARASLATGVPVT---THTAASQRDGERGRPPFLSPKLEPSRVCIGHSDDTDDLSYLTALLRGYLIG
    PVLMFMLITLLNDGCLMTIGYDHVIPS----ERP@KWNL-PVVFVS-----ASILAAVACGSSLM
                             730
                                       740
                                                            750
               720
   250
            260
                    270
                                 280
                                           290
                                                   300
   LDHIPHSAIGLEDNASASPLLGIRSV@TR----ALLIKALID@GYMK@ILVSNDWLFGFSSYVTNIMDVMDR
         L----LWIGLE----GYSSQYYENSVFHRLGLAQLPQGKLVTMMYLK-ISIS-DFLTLFSSRTGGHFFFYMP
                    770
                             780
                                 790
                                                 800
       320
                330
                          340
                                 X
   VNP--DGMAFIPLRVIPFYERRAS --HRKRC@ASL
    PSPILFCGAIISLLV---STMAASFWHKSRPDNVLTEGLAWG@TN
  820
      830
                   840 850
                                    860
LOW344-FIG1. PEP
  POLGSWNV GENOME POLYPROTEIN (CAPSID PROTEIN C; ENVELOPE GLY
 ID
     POLG#WNV
                 STANDARD; FRT; 3430 AA.
 AC
     P06935;
     01-JAN-1988 (REL. 06, CREATED)
DT
     01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT
DT
     01-MAR-1989 (REL. 10, LAST ANNOTATION UPDATE)
DE
     GENOME POLYPROTEIN (CAPSID PROTEIN C; ENVELOPE GLYCOPROTEIN M; MAJOR
     ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEINS NS1, NS2A, NS2B, NS3, NS4A,
DE
     NS4B AND NS5).
DE
08
     WEST NILE VIRUS.
OC
     VIRIDAE; SS-RNA ENVELOPED VIRUSES; FLAVIVIRIDAE.
RN
     [1] (SEQUENCE FROM N. A.)
     CASTLE E., LEIDNER U., NOWAK T., WENGLER G., WENGLER G.;
RA
RL
     VIROLOGY 149:10-26(1986).
     [2] (SEQUENCE OF 1-291 FROM N. A.)
RN
     CASTLE E. , NOWAK T. , LEIDNER U. , WENGLER G. , WENGLER G. ;
RA
RL
     VIROLOGY 145:227-236(1985)
RN
     [3] (SEQUENCE OF 255-854 FROM N. A.)
     WENGLER G. , CASTLE E. , LEIDNER U. , NOWAK T. , WENGLER G. ;
RA
     VIROLOGY 147:264-274(1985).
RL
     -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC
        LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC
CC
        PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC
        PROTEIN C AND MRNA
DR
     PIR; A25256; GNWVWV.
DR
     EMBL; M10103; FLWNVSP.
KW
     POLYPROTEIN; COAT PROTEIN; ENVELOPE PROTEIN; NONSTRUCTURAL PROTEIN;
KW
     TRANSMEMBRANE; GLYCOPROTEIN.
FT
     CHAIN
             1
                      105
                               CAPSID PROTEIN C.
FT
     PROPER
               106
                      215
FT
     CHAIN
               216
                     290
                               ENVELOPE GLYCOPROTEIN M.
FT
               291
                               MAJOR ENVELOPE PROTEIN E.
     CHAIN
                     787
FT
               788 1187
     CHAIN
                               NUNSTRUCTURAL PROTEIN NS1.
FT
     CHAIN
              1188 1354
                               NONSTRUCTURAL PROTEIN NS2A.
FT
                               NONSTRUCTURAL PROTEIN NS2B.
     CHAIN
              1355
                    1484
FT
     CHAIN
              1485
                   2109
                               NONSTRUCTURAL PROTEIN NS3.
FT
     CHAIN
              2110
                     2394
                               NONSTRUCTURAL PROTEIN NS4A.
FT
              2395
                               NONSTRUCTURAL PROTEIN NS4B.
     CHAIN
                     2579
              2580
FT
     CHAIN
                   3430
                               NONSTRUCTURAL PROTEIN NS5.
     CARBOHYD 138
FT
                     138
                               POTENTIAL.
                               POTENTIAL.
FT
               917
                      917
     CARBOHYD
```

```
962
FT
    CARBOHYD
                   562
                          POTENTIAL.
    CARBOHYD 994 994
CARBOHYD 1289 1289
CARBOHYD 1659 1659
CARBOHYD 2336 2336
FT
                          POTENTIAL.
FT
                          FOTENTIAL.
FT
                          POTENTIAL.
FT
                         POTENTIAL.
    CARBOHYD 2489 2489
CARBOHYD 2573 2573
FT
                         POTENTIAL.
                         POTENTIAL.
FT
    CARBOHYD 2739 2739
FT
                         POTENTIAL.
    CARBOHYD 2759 2759
FT
                         POTENTIAL.
    CARBOHYD 2864 2864
                          POTENTIAL.
FT
    CARBOHYD 2902 2902
FT
                         POTENTIAL.
    SEQUENCE 3430 AA; 379640 MW; 2.098737E+07 CN;
SB
Initial Score = 10 Optimized Score = 68 Significance = 3.95
Residue Identity = 21% Matches = 87 Mismatches = 231
                78 Conservative Substitutions
Gaps
                10
                                         40
                       20
                                    30
          M@TRRVVLKSAAAGTLLGGLAGCAT-----WLDRSA@AIGSIRARPITISEAGFTLTHE-
           VESHGKIGAT@AGRFSITPSAPSYTLKLGEYGEVTVDCEPRSGIDTSAYYVMS----VGEKSFLVHREW
  440
             460 470 480 490
                                 80
         60
                        70
                                        30
   ----DICGSSAG-----FLRAWPE---FFGSRKAL--AEKAVRGLRARAAGV----RTIVDVSTFDIGR
        FMDLNLPWSSAGSTTWRNRETLMEFEEPHATK@SVVALGS@EGALH@ALAGAIPVEFSSNTVKLTSGHLKCR
      510 520 530 540 550 560 570
    110 120 130 140 150 160 170
   DVSLLAEVSRAADVHIVAATGLWFDPPLSMRLRYVEELT@FFLREI@YGIEDTGIRAGIIKVATTGKATPF@
              580 590 600 610
   180 190 200 210
                                          220 230
   ELVLKAAARASLATGVP-VTTHTAASGRDGERGRPPFLS-----PKLEPSRVCIGHSDDT--
   RLV-----TVNPFVSVATANSKVLIELEPPFGDSYIVVGRGE@@INHHWHKSGSSIGKAFTTTLRGA
               650 660 670 680
 640
                                            690
         240
               250 260 270
                                      280
                                            290
   DDLSYL--TALLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSW@TRALLIKALID@GYMK@ILVSNDWLFG
   QRLAALGDTAVDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSVITQGLL-GALLLVMGINARDRSIAMTFL
       710 720 730 740 750 760 770
                      330
              320
                              340 X
   FSSYVTNIMDVMDRVNPD-GMAFIPLRVIPFYERRASHRKRC@ASL
     AVGGVLLFLSV--NVHADTGCAIDIGRØELRCGSGVFIHNDVEAWMDRYKFYPETP
       780 790 800 810 X 820
9. LOW344-FIG1. PEP
  ATXB$LEIDO PROBABLE E1-E2 TYPE CATION ATPASE 18 (EC 3.6.1.-).
ID
    ATXB$LEIDO STANDARD; PRT; 974 AA.
AC
    P12522;
    01-OCT-1989 (REL. 12, CREATED)
DT
DT
    01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT
    01-OCT-1989 (REL. 12, LAST ANNOTATION UPDATE)
    PROBABLE E1-E2 TYPE CATION ATPASE 18 (EC 3.6.1.-).
DE
08
    LEISHMANIA DONOVANI.
    EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; MASTIGOPHORA.
OC
    [1] (SEQUENCE FROM N.A.)
RN
```

```
MEADE J.C., HUDSON K.M., STRINGER S.L., STRINGER J.R.;
RA
    MOL. BIOCHEM. PARASITOL. 33:81-92(1989).
RL
    -!- CATALYTIC ACTIVITY: ATF + H(2)0 = ADP + ORTHOPHOSPHATE.
CC
CC
    -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC
        (E1-E2 ATPASES).
    -!- SIMILARITY: THE TWO L. DONOVANI CATION-TRANSPORTING ATPASE GENES
CC
CC
        ARE 98% HOMOLOGOUS.
    EMBL; J04004; LDCATP2.
DR
DR
    PROSITE; PS00154; ATPAGE_E1_E2.
    HYDROLASE; ATP HYDROLYSIS; TRANSMEMBRANE; PHOSPHORYLATION;
KW
    MAGNESIUM; ATP-BINDING.
KW
    TRANSMEM
                            PUTATIVE.
FT
              93
                   137
FT
    TRANSMEM
             118
                            PUTATIVE.
    TRANSMEM 265 286
TRANSMEM 295 321
TRANSMEM 631 651
FT
                            FUTATIVE.
FT
                            PUTATIVE.
FT
                           PUTATIVE.
FT
    TRANSMEM
             662 684
                           PUTATIVE.
    TRANSMEM 698 712 PUTATIVE.
TRANSMEM 738 761 PUTATIVE.
TRANSMEM 813 840 PUTATIVE.
TRANSMEM 869 887 PUTATIVE.
MOD_RES 351 351 PHOSPHORYLATION.
FT
FT
FT
FT
FT
    SEQUENCE 974 AA; 107304 MW; 5132373 CN;
SQ
                 18 Optimized Score = 68 Significance = 3.95
24% Matches = 93 Mismatches = 210
Initial Score
Residue Identity =
Gaps
                  82 Conservative Substitutions
                                                           0
                             20
                                     30
                                             40
                     10
           M@TRR----VVLKSAAAGTLLGGLAGCATWLDRSA@AIGSIRARPITISEAGFTLTHE---
             FLDPPRPDTKDTIRRSKEYSVDVKMITGDKLLIAKEMC-RMLDLDPNILTADKLP@IKDANDLPEDLGEKYG
                                           550
  500
          510
               520 530
                               540
                                       100
                                90
                                                110 120
              70 80
      60
   DICGSSAGFLRAWPEFFGSRKALAEKAVRGLRAR---AAGVRTIVDVSTFDIGRDVSLLAEVSRAADVHIVA
   DMMLSVGGFAQVFPE-----HKFMIVETLRQRGYTCAMTGDGVNDAPALKRADV--GIAVHGATDAARAA
                      590 600 610 620 630
 570
       580
                                     160
              140
                           150
                                             170
      130
   ATGLWFDPPLSMRLRYVE-----ELT@FFLREI@YGIEDT---GIRAGIIKVATTGKA----TPF@---E
   ADMVLTEPGLS---VVVEAMLVSREVF@RMLSFLTYRISATL@LVCFFFIACFSLTPKAYGSVDPNF@FFHL
               650 660 670 680 690 700
                    200
                             210 220 230
  180 190
                                                      240
   LVLKAAARASLATGVPVT---THTAASGRDGERGRPPFLSPKLEPSRVCIGHSDDTDDLSYLTALLRGYLIG
   PVLMFMLITLLNDGCLMTIGYDHVIPS----ERPQKWNL-PVVFVS-----ASILAAVACGSSLM
                          730 740
             720
           260 270
                             280
                                       290
                                               300
   250
   --LWIGLE---GYSSQYYENSWFHRLGLAQLPQGKLVTMMYLK-ISIS-DFLTLFSSRTGGHFFFYVP
                  7'70 780 790 800
             330
                        340
   DRVNPDGMAFIPLRVIPFYERRAS -- HRKRC@ASL
         PSPILFCGAIISLLV---STMAASFVHKSRPDNVLTEGLAVG@TN
  820
          830
                   840
                          850
```

```
ID
    ATPOSOENBI STANDARD; PRT; 511 AA.
AC
    P05492;
    01-NOV-1988 (REL. 09, CREATED)
DT
    01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
DT
    01-NOV-1988 (REL. 09, LAST ANNOTATION UPDATE)
DT
    ATP SYNTHASE ALPHA CHAIN, MITOCHONDRIAL (EC 3.6.1.34).
DE
05
    DENOTHERA BIENNIS.
OG
    MITOCHONDRION.
OC
    EUKARYOTA; PLANTA; SPERMATOPHYTA; ANGIOSPERMAE.
    [1] (SEQUENCE FROM N. A.)
RN
    SCHUSTER W. , BRENNICKE A. ;
RA
    MOL. GEN. GENET. 204:29-35(1986).
RL
    -!- FUNCTION: THIS IS ONE OF THE 5 CHAINS OF THE ENZYMATIC COMPONENT
CC
CC
       (COUPLING FACTOR CF(1)) OF THE MITOCHONDRIAL ATPASE COMPLEX.
DR
    EMBL; XO4023; MIOBATPA
DR
    PROSITE; PS00152; ATPASE ALPHA BETA.
    ATP SYNTHESIS; CF(1) COUPLING FACTOR; HYDROGEN ION TRANSPORT;
KW
    HYDROLASE; ATP-BINDING; MITOCHONDRION.
KW
    NP_BIND 171 178 ATP (BY SIMILARITY). ACT_SITE 373 373 BY HOMOLOGY.
FT
FT
    SEQUENCE 511 AA; 55596 MW; 1250759 CN;
SØ
Initial Score = 9 Optimized Score = 68 Significance = 3.95
Residue Identity = 23% Matches = 91 Mismatches = 223
Gaps = 67 Conservative Substitutions = 0
       X 10 20 30
                                     40
   MQTRRVVLKSAAAGTLLGGLAGCATVLDRSAQ---AIG-SIR----ARPITISEAGFTLTHEDICGSSAGFL
       MEFSPRAAELTTLLESRITNFYTNFØVDEIGRVISVGDGIARVYGLNEIØAGEMVEFASGVKGIAL
          10 20 30 40 50
         70
                80 90
                               100 110
                                              120
   RAWPE----FFGSRKALAEKAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAEVSRAADVHI--VAATG---
     NLENENVGIVVFGSDTAIKE---GDLVKR---TGSIVDV----PAGKSLLGRVVDALGVPIDGRGALGDHE
                                  100 110 120
    70 80
                    90
                     150 160 170 180
      130 140
   ---LWFDPPLSMRLRYVEELT@FFL-REI@YGIEDTGIRAGIIKVATTGKATPF@ELVLKAAARASLATGVP
     RRRVEVKVPGIIERKSVHEPMQTGLKAVDSLVPIGRGQRELIIGDRQTGKTAIAIDTILNQKQMNSRATSES
   130 140 150 160 170 180 190
                210 220 230 240 .
   VT---THTAAS@-RDGERGRPPFLS--PKLEPSRVCIGHSDDTDDLSYLTALLRGYLIG---LDHIPHSAI-
   ETLYCVYVAIG@KRSTVA@LV@ILSEGNALEYSILVAATASDPAPL@FL-APYSGCAMGEYFRDNGMHALII
 200 210 220 230 240 250 260
              270 280 290
                                    300 310
   YDDLSKQAVAYRGMSLLLRRFPGREA----FPGDVFYLHSRLLERAAKRSDQTGAGS--LTALPVIETQAGD
                         300 310 320 330
        280 290
 320 330 340 X
   GMAF IPLRVIPFYERRASHRKRC@ASL
```

VSAYIPTNVI-----SITDGOXCLETELFYRGIR

340

350 X 360

AIPOSUENBI AIP SYNTHASE ALFHA CHAIN, MITOCHONDRIAL (EC 3.6.1.

Results file low344-fig1.res made by maryh on Wed 17 Apr 91 12:41:07-PDT. LOW344-FIG1. SEQ Query sequence being compared: 43406 Number of sequences searched: 156 Number of scores above cutoff: Results of the initial comparison of LOW344-FIG1. SEQ with: Data bank : GenBank 65.0, all entries Data bank : UEMBL 24_65, all entries 100000-U50000-Μ В Ε F10000-S E 5000-Ø U Ε Ν С Ε S 1000-500-100-

10-

5-										
-										
_										
-										

_										
	* *					*		×	*	
0										
1 1	:::::::	;	:	<u> </u>	;	1	1	:	:	
SCORE 0:	::99:::	198	296	395	494	593	691	790	889	
STDEV -1	8									

PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0. 33		
Cutoff score	79		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to sav	/e 20	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	31	31	12.75
Times:	CPU 00:54:30,02		Total Elapsed

Number	of	residues:	54775335
Number	οf	sequences searched:	43406
Number	σf	scores above cutoff:	156

The scores below are sorted by initial score. Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Seque	nce Name	Description	Length	Init. Score	•	Sig. F	rame
		**** 67 standard deviations	above me	 ean ***	·*		
1.	PSEPTE	Plasmid pCMS1 (from P. diminuta	1322	889	1304	67. 29	0
		**** 61 standard deviations	above me	ean ***	* **		
2.	M22863	Figure 1. Nucleotide sequence	1326	809	1309	61.02	0
		**** 44 standard deviations	above me	ean ***	×		
3.	FVBOPD	Flavobacterium sp. parathion h	1693	599	1281	44.55	0
		**** 8 standard deviations	above me	ean ***	; * .		
4.	X15898	Eimeria temella mRNA for sporo	957	136	360	8. 23	0
		#*** 6 standard deviations	above me	ean ***	*		
5.	X14805	Mouse mRNA encoding DNA (cytos	4973	112	569	6. 35	0
6.	CELPOLII	C. elegens PNA pulymerase II la	12993	111	526	6. 27	0
7.	HSHEPSH	Human hepatome mRNA for serine	2363	108	562	6.04	0
8.	HUMHPSNA	Human hepsin milNA; complete cd	1783	108	456	6.04	0
		**** 5 ctandord deviations	above m	ean ***	X		
9.	BLYAMY2	Barley (H. vulgara) alpha-amyla	1588	107	590	5. 96	0

10.	BLYAMYAA	Barley alpha-amylase type A is	1588	107	592	5. 96	0
11.	BOVGABARB	Bovine mRNA for gamma-aminobut	3010	106	574	5.88	0
12.	BOVIGCAB	Bovine Ig germline gamma-2-cha	1979	104	479	5. 73	0
13.	BTIGG2HC	Bovine Ig germline heavy chain	1979	104	479	5. 73	0
14.	RRATP2	Rhodospirillum rubrum gene clu	4240	104	194	5. 73	0
15.	PDEMDH	P. denitrificans methanol dehyd	2314	103	425	5.65	0
16.	MUSGT2A	M. musculus glucose transporter	2521	102	232	5. 57	0
17.	SMASFUABC	S. marcescens periplasmic-bindi	4583	101	583	5.49	0
18.	RABIGHAB	Rabbit Ig mu chain secreted fo	1953	101	489	5. 49	0
19.	HUMIGHBD	Human Ig umproductively rearra	1127	100	384	5.41	0
20.	MUSRGEB3	Mouse 185, 5.85, 285 rRNA gene	3061	98	360	5. 25	0

The scores below are sorted by optimized score. Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

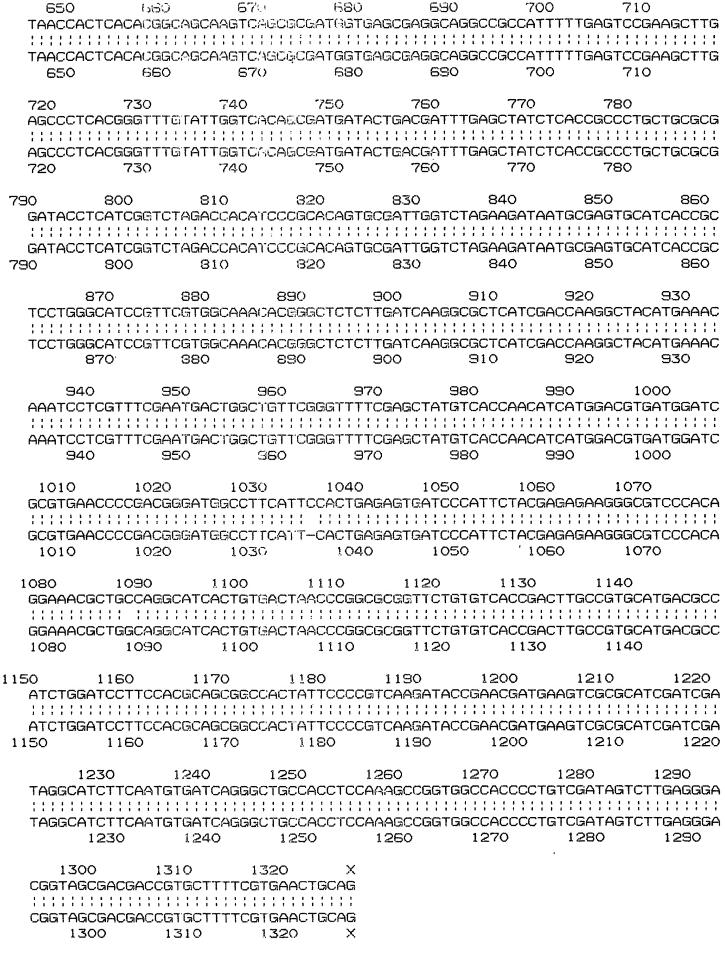
Seque	nce Name	Description	Length	Init. Score	•	Sig.	Frame
1.	M22863	**** 271 standard deviations Figure 1. Nucleotide sequence **** 269 standard deviations	1326	809	1309	271. 26	s o
2.	PSEPTE	Plasmid pCMS1 (from P. diminuta **** 261 standard deviations	1322	889	1304	269. 45	5 0
3.	FVBOPD	Flavobacterium sp. parathion h	1693	599	1281	261.10	0
4.	TRN21TNPA	Transposon Tn21 tnpA gene for **** 12 standard deviations	3176	87	606	15. 98	3 0
5.	X17379	Sorghum vulgare mRNA for phosp **** 11 standard deviations	3147	80	597	12. 71	ı o
6.	NEUTRP1	n. crassa trifunctional tryptop **** 10 standard deviations	2750	85	594	11.62	2 0
7.	BLYAMYAA	Barley alpha-amylase type A is			592	10. 89	9 0
8.	PDUMER	Plasmid pDU1358 (from S. marces		95	590	10.17	7 0
9.	BLYAMY2	Barley (H. vulgere) alpha-amyla		107	590	10.17	7 0
10.	BPECYADE	Bordetella pertussis cyaD gene		82	590	10.17	7 0
		**** 8 standard deviations	above m	ean **;	(
11.	MUSNFMG	Mouse NF-M gene for middle-mol **** 7 standard deviations	5471 above m	81 ean **:	586 **	8. 72	5 0
12.	ACFTS140A	Fujinami sarcoma virus tempera	2715	86	584	7. 99	3 0
13.	SVGSII	Streptomyces viridochromogenes		82	584	7. 99	3 0
14.	SMASFUABC	S. marcescens periplasmic-bindi	4583	101	583	7. 63	3 0
15.	HS11UL	Herpes simplex virus type 1 (H	108360	94	582	7. 26	6 0
		**** 6 standard deviations			{ 		
16.	ATUPRIREP	A. tumefaciens plasmid pRiA4b r	4638	80	581	6.90	0 0
17.	HUMASPX	Human nonerythroid alpha-spect	7787	80	581	6.90	0
18.	SERCYSA	S. erythraea rhodenese-like pro	3373	81	581	6. 90	0
19.	X51950	E. coli purHD operon for AICAR	3432	94	580	6. 54	1 0
20.	MUSHCK	Mouse hok gene for tyrosine ki	1960	81	580	6. 54	1 0

1. LOW344-FIG1. SEQ

M22863 Figure 1. Nucleotide sequence of Flavobacterium op

LOCUS M2263 1326 bp ds-DNA UNA 15-JUN-1989
DEFINITION Figure 1. Nucleotide sequence of Flavobacterium opd gene fragment.
ACCESSION M22863
KEYWORDS .
SOURCE
ORGANISM Unknown
Unclassified.
REFERENCE 1 (bases 1 to 1326)

```
AUTHORS
         Harper, L. L., McDamel, C. S., Miller, C. E. and Wild, J. R.
         Dissimilar Plasmids Isolated from Pseudomonas diminuta MG and a
  TITLE
         Flavobacterium sp. (ATCC 27551) Contain Identical opd Genes
  JOURNAL
         Appl. Environ. Microbiol. 54, 2586–2589 (1988)
  STANDARD
         unannotated staff_entry
BASE COUNT
            279 a
                   363 c
                          392 g
                                286 t
ORIGIN
Initial Score
                809
                   Optimized Score =
                                    1309
                                        Significance = 271, 26
Residue Identity =
                98%
                   Matches
                                    1316
                                        Mismatches
                                                      Ъ
                 7
Gaps
                    Conservative Substitutions
                                                      0
         10
                20
                        30
                               40
                                       50
                                              60
                                                      70
  Х
  CTGCAGCCTGACTCGGCACCAGTCGCTGCAAGCAGGTCGTAAGCAATCGCAAGGGGGCAGCATGCAAACGA
   CTGCAGCCTGACTCGCACCAGTCGCTGCAAGCAGAGTCGTAAGCAATCGCAAGGGGCCAGCATGCAAACGA
         10
                20
                        30
                               40
                                      50
                                              60
  X
                                                      70
  90
                     100
                             110
                                    120
                                            130
                 11
            160
                    170
                           180
                                   190
                                          200
     150
  ATCGATCGGCACAGGCGATCGGATCAATACGTGCGCGTCCTATCACAATCTCTGAAGCGGGTTTCACACTGA
   ATCGATCGGCACAGGCGATCGATCACTACGTGCGCGTCCTATCACAATCTCTGAAGCGGGTTTCACACTGA
     150
            160
                    1.70
                           180
                                   190
                                          200
                                                  210
                          250
                                         270
   220
           230
                  240
                                 260
                                                280
  CTCACGAGGACATCTGCGGCAGCTCGGCAGGATTCTTGCGTGCTTGGCCAGAGTTCTTCGGTAGCCGCAAAG
   CTCACGAGGACATCT-CGGCAGCTCGGCAGGATTCTTGCGTGCTTGGCCAGAGTTCTTCGGTAG-CGCAAAG
   220
           230
                   240
                           250
                                  260
                                          270
                                                  280
                        320
  290
        1,300
                                       340
                 310
                                330
                                               350
                                                      360
   CTCTAGCGĞAAAAGGCTGTGAGASSATTGCGCGCCAGAGCGGCTGCGCGTGCGAACGATTGTCGATGTCGA
   CTCTAGCGCAAAAGGCTGTGAGAGBATTBCGCGCCAGAGCGGCTGCGAACGATTGTCGATGTGTCGA
   290
           300
                  310
                          320
                                 330
                                         340
                                                350
               380
                       390
                              400
                                      410
                                             420
  CTTTCGATATCGGTCGCGACGTCAGTTTATTGGCCGAGGTTTCGCGGGGCTGCCGACGTTCATATCGTGGCGG
   CTTTCGATATCGGTCGCGACGTCAGTTTATTGGCCGAGGTTTCGCGGGCTGCCGACGTTCATATC-TGGCGG
         370
                 380
                        390
                                       410
  360
                                400
                                             . 420
      440
              450
                     460
                             470
                                    480
                                            490
                                                    500
  CGACCGGCTTGTGGTTCGACCCGCCACTTTCGATGCGATTGAGGTATGTAGAGGAACTCACAC-AGTTCTTC
   CGACCGGCTTGTGGTTCGACCCGCCAC1TTCGATGCGATTGAGGTATGTAGAGGAACTCACACTAGTTCTTC
 430
        440
                450
                        460
                               470
                                      480.
                                              490
                                                      500
              520
                     530
                             540
                                     550
                                            560
  CTGC-GTGAGATTCAATATGGCATCGAAG-ACACCGGAATTAGGGCGGGCATTATCAAGGTCGCGACCACAG
   CTGCGGTGAGAT/TCAATATGGCATUGAAGTACACCGGAATTAGGGCGGGCATTATCAAGGTCGCGACCACAG
       510
              520
                      530
                             540
                                     550
                                             560
                                                    570
      580
             590
                     600
                            610
                                   620
                                           630
                                                   640
  GCAAGGCGACCCCTTTCAGGAGTTAGTGTTAAAGGCGGCCGCCCGGGCCAGCTTGGCCACCGGTGTTCCGG
   GCAAGGCGACCCCTTTCAGGAGT (AGTGTTAAAGGCGGCCGCGGGCCAGCTTGGCCACCGGTGTTCCGG
      580
             590
                    600
                            610
                                   620
                                           630
                                                   640
```



2. LOW344-FIG1. SEQ

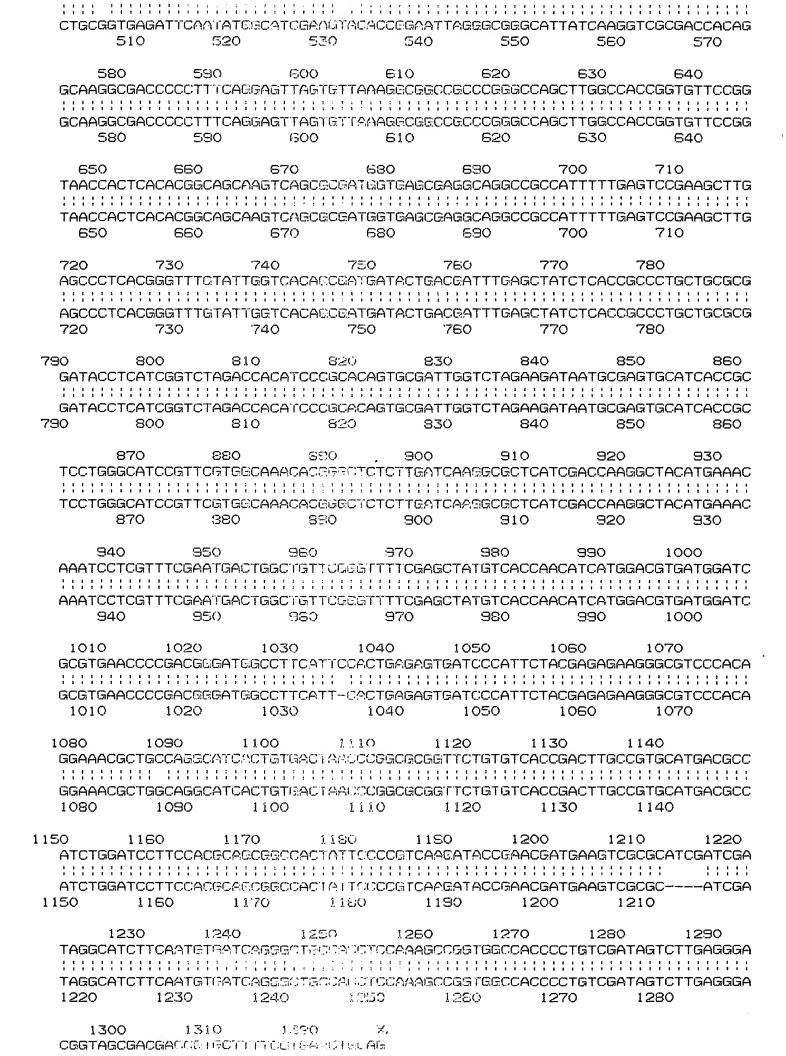
PSEPTE Plasmid pCMS1 (from P. diminuta) phosphodiesterase

LOCUS PSEPTE 1322 bp ds-DNA BCT 15-MAR-1989 DEFINITION Plasmid pCMS1 (from P. diminuta) phosphodiesterase (opd) gene:

```
M20392
ACCESSION
KEYWORDS
          phosphotriesterase.
SOURCE
          Plasmid pCMS1 (from Pseudomonas diminuta) DNA.
  ORGANISM
          Plasmid pCMS1
          Prokaryota; Bacteria; Plasmid pCMS1.
          1 (bases 1 to 1322)
REFERENCE
  AUTHORS
          McDaniel, C. S., Harper, L. L. and Wild, J. R.
  TITLE
          Cloning and sequencing of a plasmid-borne gene (opd) encoding a
          phosphotriesterase
          J. Bacteriol. 170, 2306-2311 (1988)
  JOURNAL
  STANDARD
          simple staff review
FEATURES
                 Location/Qualifiers
    CDS
                 63.. 1040
                 /note="phosphotriesterase protein"
BASE COUNT
                           392 g
                                  285 t
             278 a
                    367 c
ORIGIN
          5 bp upstream from PstI site.
Initial Score
                 889
                    Optimized Score =
                                     1304
                                         Significance = 269.45
                 98%
                                     1313
                                          Mismatches
Residue Identity =
                    Matches
                                                        6
Gaps
                 11
                     Conservative Substitutions
                                                        0
         10
                 20
                        30
                                40
                                        50
                                                60
                                                        70
   CTGCAGCCTGACTCGCCAGTCGCTGCAAGCAGAGTCGTAAGCAATCGCAAGGGGGCAGCATGCAAACGA
   CTGCAGCCTGACTCGCCACCAGTCGCTGCAAGCAGAGTCGTAAGCAATCGCAAGGGGGCAGCATGCAAACGA
                                40
                                        50
                                                60
   X
         10
                 20
                        30
                                                        70
               90
                                      120
                                              130
       80
                      100
                              110
   80
               90
                      100
                              110
                                      120
                                              130
                                                     140
     150
             160
                    170
                            180
                                    190
                                            200
                                                    210
   ATCGATCGGCACAGGCGATCGGATCACTACGTGCGCGTCCTATCACAATCTCTGAAGCGGGTTTTCACACTGA
   ATCGATCGGCACAGGCGATGCGATCAATACGTGCGCGTCCTATCACAATCTCTGAAGCGGGTTTCACACTGA
     150
             160
                     170
                            180
                                    190
                                            200
                                                    210
                           250
                                                  280
           230
                   240
                                   260
                                          270
   CTCACGAGGACATCTGCGGCAGCTCGGCAGGATTCTTGCGTGCCTGGCCAGAGTTCTTCGGTAGCCGCAAAG
   CTCACGAGGACATCT-CGGCAGCTCGGCAGGATTCTTGCGTGCTTGGCCAGAGTTCTTCGGTAG-CGCAAAG
           230
                    240
                            250
                                   260
                                           270
                                                    280
   220
  290
          300
                         320
                                 330
                                         340
                                                350
                 310
                                                        360
   CTCTAGCGGAAAAGGCTGTGAGAGGATTGCGCGCCAGAGCGGCTGCGAACGATTGTCGATGTCGA
   CTCTAGCGGAAAAGGCTGTGAGAGGATTGCGCGCCCAGAGCGGCTGCGCGAACGATTGTCGATGTCGA
   290
           300
                   310
                           320
                                  330
                                          340
                                                  350
                380
                        390
                               400
                                       410
                                               420
                                                       430
   CTTTCGATATCGGTCGCGACGTCAGTTTATTGGCCGAGGTTTCGCGGGGCTGCCGACGTTCATATCGTGGCGG
   CTTTCGATATCGGTCGCGACGTCAGTTTATTGGCCGAGGTTTCGCGGGGCTGCCGACGTTCATATC-TGGCGG
  360
          370
                 380
                         390
                                 400
                                         410
                                                 420
       440
              450
                      460
                              470
                                      480
                                             490
                                                      500
   CGACCGGCTTGTGGTTCGACCCGCCACTTTCGATGCGATTGAGGTATGTAGAGGAACTCACAC-AGTTCTTC
   CGACCGGCTTGTGGTTCGACCCGCCACTTTCGATGCGATTGAGGTATGTAGAGGAACTCACACTAGTTCTTC
 430
         440
                 450
                        460
                                470
                                        480
                                                490
                                                        500
      510
              520
                      530
                               540
                                      550
                                              560
                                                      570
```

CTGC-GTGAGATTCAATATGGCATCGAAG-ACACCGGAATTAGGGCGGGCATTATCAAGGTCGCGACCACAG

complete cds.

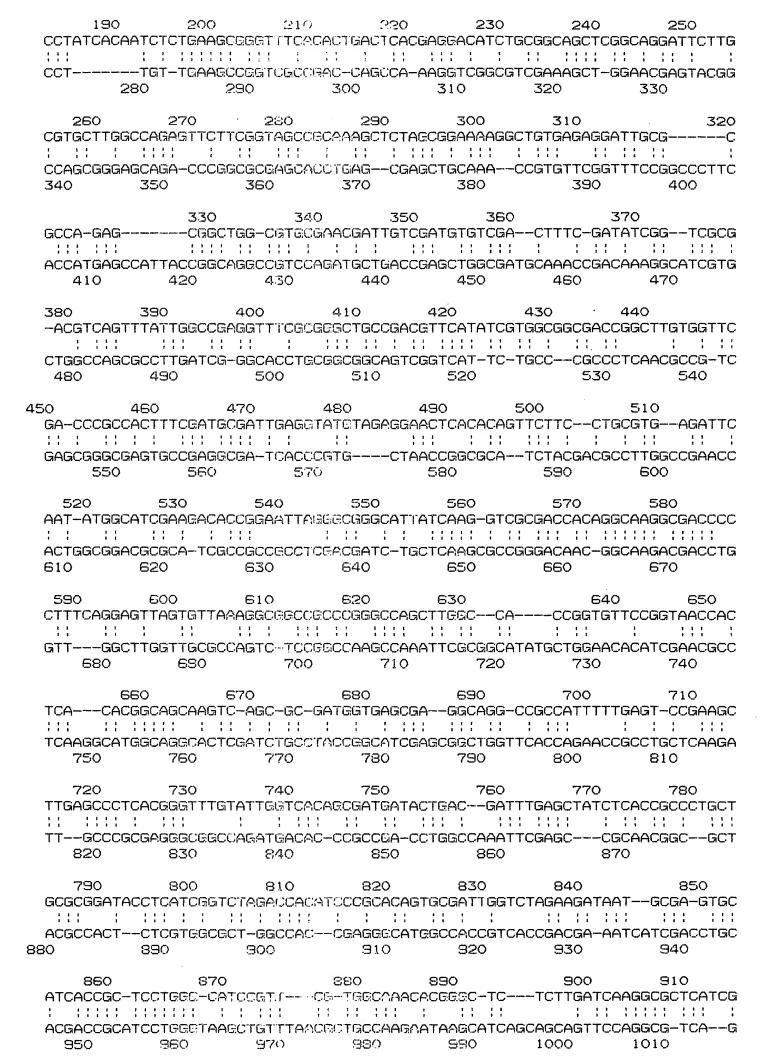


```
CGGTAGCGACGACCGTGCTTTTCCGTGATETISCAG
 1290
          1300
                   1310
                           1330 X
3. LOW344-FIG1. SEQ
  FVBOPD
            Flavobacterium sp. parathion hydrolase gene, compl
LOCUS
                      1803 bp ds-DNA
                                              BCT
                                                       15-JUN-1990
           Flavobacterium sp. parathion hydrolase gene, complete cds.
DEFINITION
ACCESSION
           M29593
KEYWORDS
           parathion hydrolase.
SOURCE
           Flavobacterium sp. (strain ATCC 27551) DNA, clone pPDL2.
  ORGANISM
           Flavobacterium sp.
           Prokaryota; Bacteria; Gracilicutes; Scotobacteria; Neisseriaceae;
           Flavobacterium; sp.
REFERENCE
           1 (bases 1 to 1693)
  AUTHORS
           Mulbry, W. W. and Karns, J. S.
  TITLE
           Parathion hydrolase specified by the Flavobacterium opd gene:
           Relationship between the gene and protein
  JOURNAL
           J. Bacteriol. 171, 6740-6746 (1989)
           simple staff_entry
  STANDARD
FEATURES
                   Location/Qualifiers
     misc signal
                   312..317
                   /note="-35 region"
     misc_signal
                   334. . 339
                   /note="-10 region"
                   408..411
     misc_binding
                   /nute="ribosme binding site"
     CDS
                   419. . 1515
                   /note="parathion hydrolase"
                              477 g
BASE COUNT
              372 a
                      457 c
                                      347 t
ORIGIN
           1 bp upstream of BamHI site.
                                               Significance = 261.10
Initial Score
                   599 Optimized Score
                                          1281
                                      =
Residue Identity =
                                          1304
                                                               14
                   SEX Mortables
                                               Mismatches
Gaps
                    28
                       Conservative Substitutions
                                                                0
                                             40
                                                      50
                   10
                            20
                                    30
                                                               60
            CTGCAGCCTGACTCGCACCAGTCGCTGCAAGCAGTCGTAAGCAATCGCAAGGGGGCAGC
            CGGTTCAGATCTGCAGCCTGACTCGCCACCAGTCGCTGCAAGCAGAGTCGTAAGCAATCGCAAGGGGGCCAGC
    350
             360
                     370
                              380
                                       390
                                                400
                                                         410
        70
                 80
                          SO
                                     100
                                             110
                                                      120
                                                               130
   ATGCAAACGAGAAGGGTTGTGCTCAAGTCTGCG---GCCGCAGGAACTCTGCTGGGCGGCCTGGCTGGTGC
   ATGCAAACGAGAAGGETTGTGCTCAAGTCTECEGCCGCCGCAGGAACTCTGCTCGGCGGCCTGGCTGGTGC
  420
           430
                    440
                            450
                                     460
                                              470
                                                       480
                                                                490
         140
                  150
                           180
                                    170
                                             180
                                                       190
   GCGA-CGTGGCTGGATCGATCGGCACGCACGCATCGGATCAATA-CGTGCGC-GTCCTATCACAATCTCTGAA
   GCGAGCGTGGCTGGATCGATCGCACACACGCGATCGCATCACATACCGTGCGCGGTCCTATCACAATCTCTGAA
                                                              560
         500
                  510
                           520
                                   530
                                            540
                                                     550
         210
                  220
                                            250
                           230
                                    240
                                                     260
                                                              270
   GCGGGTTTCACACTGACTCACGAGGACATCTGCGGCAGCTCGGCAGGATTCTTGCGTGCTTGGCCAGAGTTC
   GCGGGTTTCACACTGACTCACGACCACCACCACGCAGCAGCAGCAGCAGCATTCTTGCGTGCTTGGCCAGAGTTC
       570
                580
                         590
                                  EOO
                                          610
                                                   620
                                                            630
       280
                290
                         300
                                  310
                                             320
                                                      330
                                                               340
   TTCGGTAGCCGCAAAGCTCTAGCGHAAAAGGCTGTGAGAGGATTG----CGCGCCAGAGCGGCTGGCGTGCGA
```

TTCGGTAGCCGCAAAGCTCTAGCGT44446EEEC FGTGAGACGATTGCGCCGCGCCAGAGCGGCTGGCGTGCGA

640	65O	ميدين البيا	5 70	680	690	700
35	50 3Š0	370	380	390	400	410
ACGATTGT	CGATGTGTCGA		· · · · · · · · · · · · · · · · · · ·			CGCGGGCTGCC
	CGATGTGTCGAC	UTTTOMATATO	GETCGCGACG	TCAGITTTATT	GGCCGAGGTTT	CGCGGGCTGCC
710	720	730	740	750	760	770
420	430	440	450	460	470	480
GACGIICA	TATCGTGGCGG(GATGCGATTGA	
	TATOGTGGCGG					
780	790 (300 8	310 E	320 8	30 84	.0 850
490 GAACTCAC	500 ACAGTTCTTCC	510 TYAZANTADAN	520 CASTATIGGOS	530 TCGAAGACAC	540 CGGAATTAGGG	550 CGGGCATTATC
11111111				1::::::::::	:::::::::::::::::::::::::::::::::::::::	
	ACAGTTCTTCC 860 870					CGGGCATTATC 920
560	570	580	580	600	610 6	i20
	:GACCACAGGCA/					
	::::::::::::::::::::::::::::::::::::::					
930		950	960	970	980	990
630	640 6 5	50 6E	:0 67	70 68	0 690	700
	TGTTCCGGTAA					
	::::::::::::::::::::::::::::::::::::::					GGCCGCCATTT
1000	1010	1020	1030	1040	1050	1060
71		30 73				, . .
TTGAGTCC	GAA-GCTTGAG 					TTTGAGCTATC
	GAAGGCTTGAGG					
1070	1080	1090	1100	1110	1120	1130
78 TCACCGCC	o CT-GCTGCG0				20	
1 1 1 1 1 1 1 1			1 1 1 1 1 1 1 1 1 1 1 1 1 1	:::::::::::::::::::::::::::::::::::::::	:::::::::::::::	
TCACCGCC	CTCGCTGCGCG(1150	CEGATY\CC\CA -1160	TCGGTCTAGA 1170			TGGTCTAGAAG 200
	:50 S60	0.20	880	000	900	010
	iagtgcatcacci					910 CAAGGCGCTCA
ATAATECE	::::::::::::::::::::::::::::::::::::::	TITLLITE TO THE CO	TOOGTTOGTE	::::::::::::::::::::::::::::::::::::::	:::::::::: GCTCTCTTGAT	CAAGGCGCTCA
1210				250 12		
920	930	940	950	960	970	980
TCGACCAA	GGCTACATGAA	NCARATESTUR	TTTCGAATGA	CTGGCTGTTC	GGGTTTTCGAG	CTATGTCACCA
TCGACCAP	GGCTACATGAA	ACAAATCCTCG	TTYCGAATGA	CTGGCTGTTC	GGGTTTTCGAG	CTATGTCACCA
12	290 1300	1310	1320	1330	1340	1350
990	1000		1020	1030	1040	1050
ACATCATE	GACGTGATGGAT 	FOGUG FGAACC	JUGACEGGAT	ISGCCTTCATT	CCACTGAGAGT	GATUCCATE-C
ACATCATG 1360	GACGTGATGGA	CGCETEAACC	CCGACGGGAT	GGCCTTCATT	CCACTGAGAGT	GATCCCATTCC
						- · ·
1060 TACGAGAG	1070 AAGGGCGTCCC	1080 NCAGGAAACGC	1090 TECCABECAT	1100 CACTGTGACT	1110 AACCCGGCGCG	1120 GTTCTGTGTCA
1111111	1111111111				11111111111	11111 11111
TACGAGAG	AAGGGCGTCCC	中,69000000000000000000000000000000000000	TUBUALIBOAT	CACTGTGACT	AAUUUGGGG	GITCI-IGICA

1130
CCGACCTTGCGGGCGT—CATGACGCCATCTGGATCCTTCCAGCCAGCGGCCACTATTCCCCGTCAAGATACG 1500 1510 1520 1530 1540 1550 1560 1200 1210 1220 1230 1240 1250 1260 GAACGATGAAGTCGCGCATCGATAGGCATCTTCAATGTGATCAGGGCTGCCACCTCCAAAGCCGGTGGGAACGATGAAGTCGCGCATCGATCG
CCGACCTTGCGGGCGT-CATGACGCCATCTGGATCCTTCCAGCCAGCGGCCACTATTCCCCGTCAAGATACG 1500 1510 1520 1530 1540 1550 1560 1200 1200 1210 1220 1230 1240 1250 1260 GAACGATGAAGTCGCACCTCGATCGATAGGCATCTTCAATGTGATCAGGGCTGCCACCTCCAAAGCCGGTGGGAACGATGAAGTCGCGCATCGATCG
1500 1510 1520 1530 1540 1550 1560 1200 1210 1220 1230 1240 1250 1260 GAACGATGAAGTCGCGCATCGATCGATAGGCATCTTCAATGTGATCAGGGCTGCCACCTCCAAAGCCGGTGC !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
GAACGATGAAGTCGCGCATCGATAGGCATCTTCAATGTGATCAGGGCTGCCACCTCCAAAGCCGGTG ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
1270 1280 1290 1300 1310 1320 X
CCACCCTGTCGATAGTCTTGAGGGACGGTAGCGACGACCGTGCTTTTCGTGAACTGCAG
4. LOW344-FIG1. SEQ TRN21TNPA Transposon Tn21 tnpA gene for transposase.
LOCUS TRN21TNPA 3176 bp ds-DNA BCT 15-MAR-1988 DEFINITION Transposon Tn21 tnpA gene for transposase. ACCESSION X04891 KEYWORDS tnpA gene; tnpR gene; transposase; transposon. SOURCE Transposon Tn21.
ORGANISM Transposon Tn21 Prokaryota; Bacteria; Transposon Tn21.
REFERENCE 1 (bases 1 to 3176) AUTHORS Ward, E. and Grinsted, J. TITLE The nucleotide sequence of the tnpA gene of Tn21
JOURNAL Nucleic Acids Res. 15, 1799-1806 (1987) STANDARD simple automatic COMMENT [1] enum. 1 to 3:76. FEATURES Location/Qualifiers
CDS < 154
/note="tnpR gene" CDS 573023
/note="transposase (AA 1-988)" BASE COUNT 652 a 1057 c 915 g 552 t ORIGIN
Initial Score = 87 Optimized Score = 606 Significance = 15.98 Residue Identity = 52% Matches = 761 Mismatches = 472 Gaps = 212 Conservative Substitutions = 0
X 10 20 30 40 50 CTGCAGCCTGACTCGGCACCAG-TCGCTGCAAGCAGAGTCGTAAGCAATCGCAA
; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;
TCCATCCTGTCCGCCGCC-GAGCGGGAAAGCCTGCTGGCGTTGCCGGACTCCAAGGACG-ACCTGATC-CGA
60 70 80 90 100 110 GGGGGCAGCATGCAAACGAGA-AGGGTTGT-GCTCAAGTCTGCGGC-CGCAGGAACTCTGCTGGGC
CATTACA—CATTC—AACGATACCGACCTCTCGATCATCCGACAGCGGGCGGGCCAGCCA
140 150 160 170 180 190 200
120 130 140 150 160 170 180
GCCTGGCTGGGTGCGCGACGTG-GCTGGATCGATCGGCACAGGCGATCGGATCAA-TA-CGTGCGCGT
TCGCGG-TGCAGCTCTGTTACCTGCGCTTTCCCG-GCGTCATCCTGGGCGTC-GATGAACTACCGTTTCCGC
210 220 230 240 250 260 270



```
940
                              950
                                        960
   ACCAAGGCTACATGAAACAAATCCTCGTTTCGAATG~ACTGGCTGTTCGGGGTTTTCGAGCT~AT~GTCACCA
    GC-AAGGC--CAT----CAACGCCAAGGTACGTCTGTACGGGGCGCATCGG--TCAGGCGCTGATCGACGCCA
      1020
                   1030
                            1040
                                    1050
                                               1060
                                                       1070
       990
               1000
                                 1020
                                           1030
                         1010
   A-CA-TCATGGACGTGATG-GATCG-CGTGACCCCGACGGGATGGCCT-TCATTCC---ACTGAGAGTGAT
   AGCAATCA-GGCCGCGATGCGTTTGCCGCCATCGAGGCCGTCATGTCTTGGGGATTCCTTTGCCGAGAGCG-T
            1090
                    1100
                             1110
                                     1120
                                              1130
 1050
          1060
                   1070
                            1080
                                    1090
                                             1100
                                                      1110
   CCCATTCTACGAG-AGAAGGGCGTCCCACAGGAAACGCTGCCAGGCATCACTGTGACTAACCCGGCGCGCTT
   CAC---CGAGGCGCAGAAGCTCGCGCAACCCG-GTGGCTTC--GGTTTC-CTGCATCGCA-TCGGCG-AGAG
  1150
             1160
                     1170
                               1180
                                         1190
                                                   1200
                                                             1210
1120
                    1140
                              1150
                                                 1170
         1130
                                        1160
   CTGTGTCACCGACT-TGCCG---TGCATGACGCCAT--CTGGATCCTTC-CACGCAGCGGCCACTATTCCCC
   CTACGCCACC--CTGCGCCGCTATGCA--CCGGAATTCCTTG-CCGTGCTCAAGCTGCGGGCCGCCCGCC
                  1230
                            1240
                                      1250
                                               1260
     1190
             1200
                      1210
                               1220
                                         1230
                                                  1240
                                                           1250
   GTCAAGATACCGAACGATGAAGTCGCGCATC-GATCG--ATAGGCATCTTCAATGTGATCAGGGCTGCCACC
   GCCAAAAACGTGCTTGATGCCATTGAGGTGCTGCGCGGCATGAACACCGACAACGCCCGCA-AGCTGCCAGC
  1280
           1290
                   1300
                            1310
                                    1320
                                             1330
                                                      1340
                   1270
                              1280
                                       1230
                                                1300
    ---TCCAAAG-CCGG---TGGCCA---CCCCTGTCG-ATAGTCTTGAGGGACGGTAGCGACGACCGTGCTT--T
      CGATGCACCGACCGCTTCATCAAGCCGCGCTGGCAGAAACT--GGTGATG--ACCGACG-CCG-GCATCGA
          1360
                  1370
                           1380
                                   1390
                                               1400
 1350
                                                          1410
      1320
               Х
   TCGTG--AACTGCAG
    1 1 1
         1111
   CCGGCCTACTACGAACTGTGCGCG
     1420
             1430
5. LOW344-FIG1. SEQ
  X17379
            Sorghum vulgare mRNA for phosphoenolpyruvate invol
                                                      15-SEP-1990
LOCUS
           X17379
                                             UNA
                      3147 bb
           Sorghum vulgare mRNA for phosphoenolpyruvate involved in C4
DEFINITION
           photosynthesis (EC 4.1.1.31).
ACCESSION
           X17379
KEYWORDS
SOURCE
  ORGANISM Unknown
           Unclassified.
           1 (bases 1 to 3147)
REFERENCE
           Cretin, C., Keryer, E., Tagu, D., Lepiniec, L., Vidal, J. and Gadal, P.
  AUTHORS
  TITLE
           Complete cDNA sequence of sorghum phosphoenolpyruvate carboxylase
           involved in C4 photosynthesis
          Nucleic Acids Pes. 18, 658-658 (1990)
  JOURNAL
           unannotated staff entry
  STANDARD
REFERENCE
           2 (bases 1 to 3147)
           Cretin, C. D.
  AUTHORS
  JOURNAL
          Unpublished (1989) see COMMENT for author address
  STANDARD
           unannotated staff_entry
```

675 a

BASE COUNT

946 C

915 g

611 t

Initial Residue Gaps		ity :		80 51% 203	Optim Match Conse	es		:-7	740	Mis	nific match		= 12 = =	2. 71 490 0
		X CTG	C-AGC	10 -CTGA			AG1'CG	CT'——G				AGCAA	50 TCG-	-CAAGG
GTC(1460	GGTGA 14			; ; GCCCA 480	ACTGCT	!	CGGAC				GAGAT		;; GACG	:: : : CATCG 1530
GGG(1 111	;;	:	; ;	! !	GTGCT	1 1 1	TCTGC ;;;	; ; ; ;	CAGGA ; ;	AC-TO	1 1	;	120 GGCCTG :; ; FGCACG
300		40		550		560		570		580		590		. 0.0. 100.
* ! * *	: GCCCTC	1 1 1	:::::: CGT-G	CTGGA ;;;;; CTCG-	11:	CG-GC CGAGC	ACAGG ; ;	CGATC	; ; GC	AATAC ; ;	GTGCG GTG-G	:::	CTATO	190 CACAAT ; ; AGCGGT 1660
t 1 1 t	1 1	GCGG GCCG	: :	CGA	::::::::::::::::::::::::::::::::::::::	ACGAG ;	1 1 1	CTGC-	-GGCAG :::: :::	: :::	CAGGA ; ; GCCG	: ; -TCCG	: ; : :	CGTG ; GAAGCT
1 1 1	GGCCAG 	:	:: : CTG	:::	111	1 1	TCTAG	IIIII CGGCA	AAGGC ; ;; A——GC	: :	AGAGGA ; ;; AGGTGA	ATTGC	; ;	CAGAGC ; ; TACTCC 1790
: :	rggcgt : ;	1	1::	; ; ;	;;;	1 1	ACTTT	: :	TCGGT	-CGCG	1 1	;	;	raacca !!!!!
GAC*	rccggc 1800		ACGCC 181		CCTGT 182							166A60 1850		TGGCCA 1860
AGG ⁻ ;;;) 1	GGGC ::: -GGC	TGCCG ; -CAAG	ACGTT AAGT-	CATAT	CGTGG ; ; ; ; CGTGA	CGGCG ;;;; AGGTG	ACCGG 	CTTGT	GGTTC : TCCAC	GACC- : :: GGCCE	-CG	: :	SO CTTTCG ! !!! CCGTCG 1920
ATG(: CAGGGG	AGGT ; ; CGTT	ATGTA GGCCC	GAGGA GACGC	AACTCA CAC-CT	CACAG 	TTCTT : : : TCCGT	CCT ;; GCCCA	G-CGT ;;; GCCGC	GAGAT ;; CGGAC	: : : ACCAT	TATO ; ; TCAACO	:: aggto	CCATCC
570														1990
AAGA ;	;; TGAC	GAAT ; GGTG	TAGGG CAGGG	CGGGC ;;;; CGAG-	TATTAC ;;;;	CAAGG CGAGT	TCGCG ;; ;;	ACCAC ; TTC	AGGCA ;;	AGGCG AGGAG	: :: BAACCT	сст ; гатаст	::::	AGGAGT AGTCTC
	00	:	610		620		(630		6 40		650		660 CACGG
: TGC	:: ::	: : TCAC	::: :	III; CCACE	: ::	IIIII GCACG	: :: CCATG	:: CACCC	:::: GCCGG	: ; ; STCTCT	::::	: -AGTG0	; ;	

6.40		1590	700	710	720	730
	GCGATGGTGAGCG4					CGGGTT
	GTCGTCGCCAAGG/	NGO-AGTACGT	CGTCGTCGTC	AAGGAGCCGC		GTACTT
2130 214	10 2150	2160	2170	2180	2190	
740	750	760	770	780		_
TGTATTGGTCAC	AGCGATGATACTG-			-TCA-CCGCC(IGATACC
CAGATCGGCTAC	CCCTGAGACTGA	AGTACGGGAAG	iATGAAC-ATC	GGCAGCAGGC	CAGCCAAGAG	GAAGCC
2200 2210	2220	2230	2240	2250	2260	
800	810 82		830	840	850	
	AGACCACATCCCGC	CACAGTGCGAT				
	TCACCACCCTGC					
2270 22	2290) 2	300 2	310 23	320	2330
	370 880 CATCCGTTCGTGG0	068			920	
HUUSU TUU TOOU					JATUGALUAA !!!!!!!	
ACCTCC	CCGTGTGGC					GGAC
	2340 2	2350 2	360	2370	2380	
		960	970		990	
HIGHHACHAHIC	CTCGTTTCGAATGA	AC15551511C		IJCIAIIJICACI	JAACATCATU 	HALLATIA
ATCAAGAACTTO	CAGGTCCTCAAGGA		AACGAGTGGC	CATTCTTCAG	GTCACCCTG	GACCTG
2390 2400	2410	2420	2430	2440	2450	
1000 1010		1030		105		1060
ATGGATCGCGTG	AACCCCGACGGAT		TCCACTGAGA ;;;;;			ACGAGA
	TTCGCC-AAGGGAG	SACCCTESCAT	TGC-CGGCTT			CCGAG-
2460 247	°0 2480	2490	2500	2510	2520	
1070			100		1120	1130
	ACAGGAAACGCTGC					
GAACTCAAGCCC	TTTGGGAA-GCAG					
2530 25	540 2550	2560	2570	2580	25	90
	1150					
	ACGCCA-TCT					-CAAGA !!!
TCGCTGGGCACA	RAGGACATTCTTGAA	AGGCGATCCAT	TCCTGAAGCA	GGGGCTGCGT	CTGCGCAATC	CCTACA
2600	2610 262	20 263	so 264	0 265	D. 26 E	Ю
1200		1220			1250	
	GAAGTCGCGCATCG					
	GAACGTG-TTCC					
2670	2680	2890	2700	2710	2720	2730
	1280					
	ACCCCTGTCGATAG1 					AACTGC
ACGCCGCAGCCG	CCGCTGTCCAAGGA	AGTICGCCGAC	GAGAACAAGC	CCGCCGGACT	GGTGAAGCTG	AACGGC
2740	2750	2760	2770	2780	2790	2800
×						
-AG ::					,	

GAGCGAGTACCGC X 2810

```
6. LOW344-FIG1. SEQ
  NEUTRP1
             n crassa trifunctional tryptophan biosynthesis gen
LOCUS
           NEUTRP1
                                 DNA
                       2750 bp
                                                PLN
                                                         01-AUG-1983
           n crassa trifunctional tryptophan biosynthesis gene trp-1.
DEFINITION
ACCESSION
           J01252
KEYWORDS
           multifunctional enzyme.
SOURCE
           neurospora crassa.
  ORGANISM
           Neurospora crassa
           Eukaryota; Plantae; Thallobionta; Eumycota; Ascomycotina;
           Pyrenomycetes; Sordariales; Sordariaceae; Neurospora; crassa.
           1 (bases 1 to 2750)
REFERENCE
  AUTHORS
           Schechtman, M. G. and Yanofsky, C.
           structure of the trifunctional trp-1 gene from neurospora crassa
  TITLE
           and its aberrant expression in escherichia coli
  JOURNAL
           J. Mol. Appl. Genet. 2, 83-99 (1983)
  STANDARD
           simple staff_review
BASE COUNT
               572 a
                                       630 t
                       821 c
                               727 g
ORIGIN
                                            594
Initial Score
               ---
                    85
                       Optimized Score =
                                                 Significance = 11.62
                                                Mismatches
Residue Identity =
                   51%
                       Matches
                                            738
                   213
                        Conservative Substitutions
                                                                  0
Gaps
                    10
                             20
                                       30
                                                40
                                                         50
            CTGCAGCCTGACTCGGC-ACCAGTCGCTGCAAGCAGAGTCGTAAGCAATCG--CAAGGGGGC
              1 1 1
   ATCACAATGTCGTCCTCAGTCGTCGACCACTCTCCCCA---CGATTC-CGCTCCTTCGCCCCTGGTGCC
    260
             270
                      280
                               290
                                           300
                                                     310
                                                              320
  60
              70
                       80
                                90
                                            100
   A---GCATGCAAACGAGAGGGTTGTGCTCAAGTCTG---CG-GCCGCAGGAAC-TCTGCTGG----GCGGC
                   AACCGCCTCCAACC--TCATCCTCATCGACAACTATGATTCGTTTACCTGGAACGTCTACCAGTACCTCGTC
                 340
      330
                          350
                                   360
                                            370
                                                     380
                                                              390
 120
          130
                    140
                             150
                                      160
                                               170
                                                            180
   CTGGCTGGGTGCGCGACGTGGCTCGATCGGCACAGGCGATCGGATCA--AT--ACGTGCGCGTCCTAT
   CTCG-AGGGCGC-CAAGGTGACCG--TC-TTCCGCA---ACGACCACATCACCATCGACGAGCTCATCGCA-
       400
                 410
                            420
                                        430
                                                 440
                                                           450
   190
             200
                       210
                                 220
                                          230
                                                   240
   CACAATCTC-TGAAGCGGGT-TTCA-CACTGACTCACGAGGACATCTGCGGCAGCTCGGCAGGATTCTTGCG
              AAGAACCCCACCCAGCTCGTCATCAGCCCTG-GGCCCG--GTCATC-CCGGCACCGACTCCGGTATCTCGCG
   460
            470
                     480
                               490
                                          500
                                                    510
                                                             520
               270
                        280
                                 290
                                          300
    260
                                                   310
   TGCTTGGC--CAGAGTTCTTCGGTAGCCGCAAAGCTCTAGCGGAAAAGGCTGTGAGAGGATTGCGCGCCAGA
    CG-ATGCCATCAG-GCACTTC----GCCGGCAAGATC-CCCATCTTTGGC-GT--GTGCATGGGCCAGCAGT
      530
                540
                            550
                                       580
                                                570
    330
               340
                        350
                                 360
                                          370
                                                   380
                                                            390.
   GCGGC--TGGCGTGCGAACGATTETCGATGTCGTCGACTTTCGATATCGGTCGCGACGTCAGTTTATTGGCCG
                       1 111 1111
                                   : : : : :
                                           GCATCTTTGACGT-CTATGG--CGGCGACGTGT--GCT-TCG---CCGGT-GAGA-TTC--TGCACGGAAAG
   590
            600
                        610
                                   620
                                                 630
                                                            640
                      420
                               4.30
                                        440
                                                  450
             410
                                                           460
   AGGTTTCGCGGGCTGCCGACGTTCATATCGTGGCGGCGACCGGCTTGTGGTTCGACCCGCCACTTTC--GAT
             11 111 111
                        1 1 11
                                   2 1 2 1
                                           1 1 1
   ACCTCTCCTGC-GCC-ACGACGGCAAGGGCGCATATGCCGGTCTGTCTCAGGATCTGCCAGTGACGAGAT
```

680

670

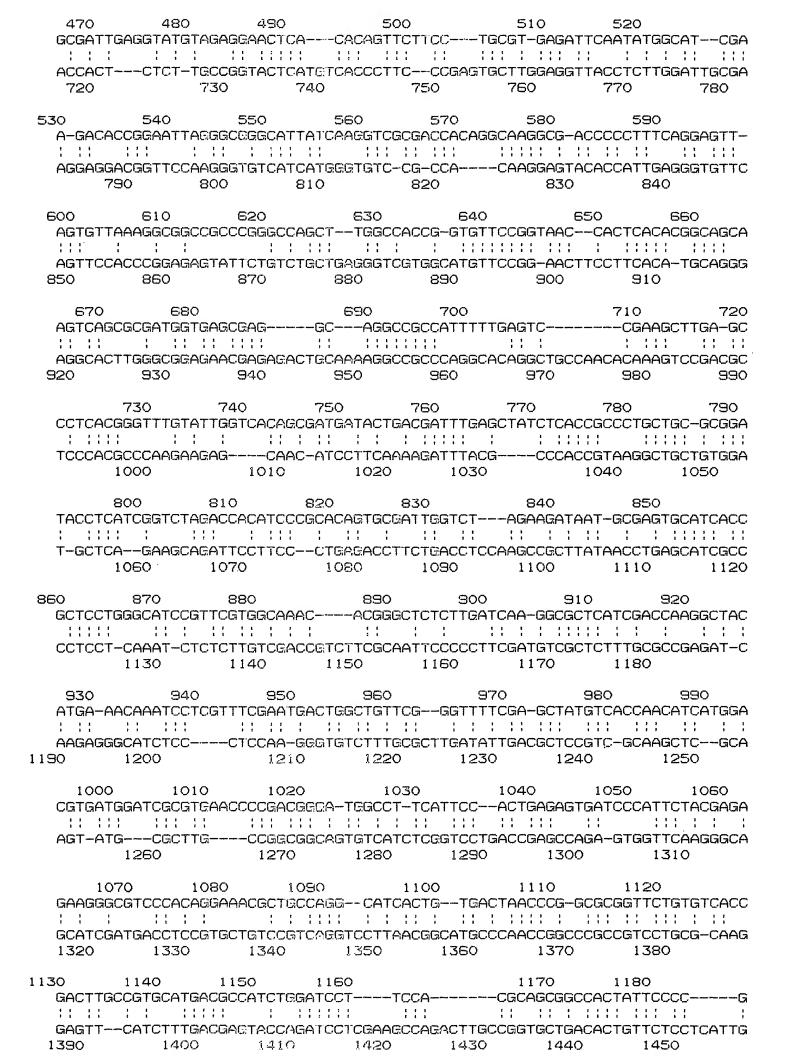
690

700

710

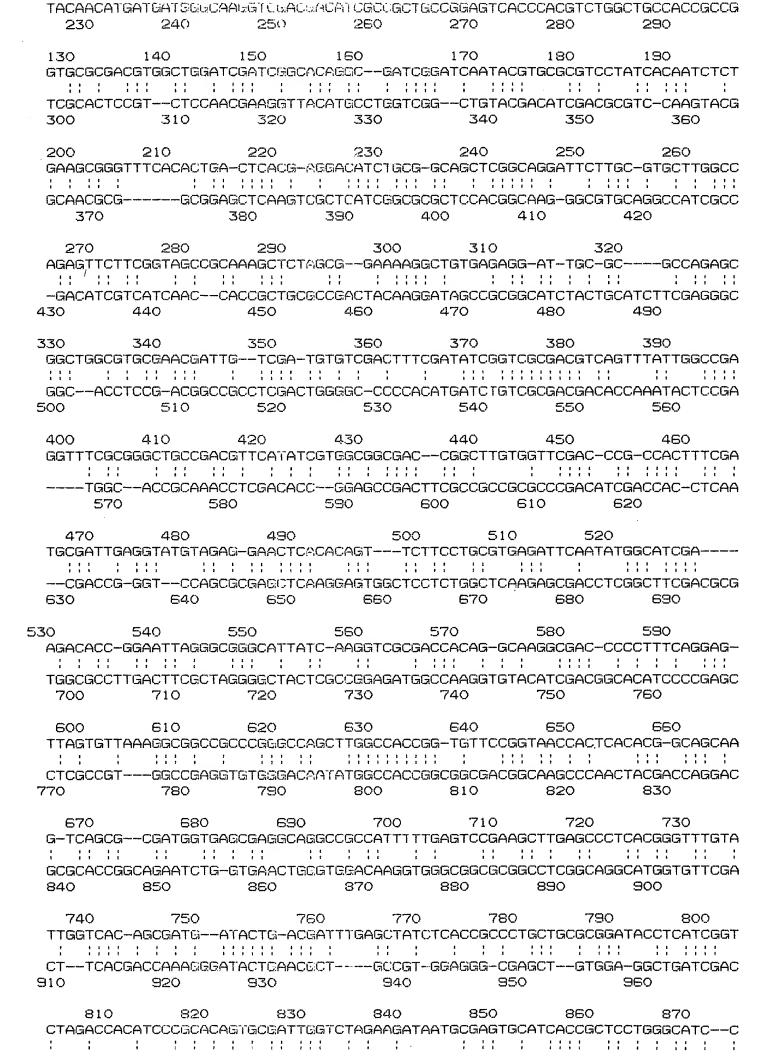
6E0

650



```
1200
                        1210
                                  1220
                                               1230
                                                         1240
   TCAAGATACCGAACGATGAAGTCGCGCATCGATCG-ATAGGCA--TCTTCAATGTGATCAGGGCTGCCACCT
              1 1 1 1 1
                          1 ;
                                               1 1 1
                                                      1 1 1
                                                           1 1
   TCAAGATGCTCGAGTATGA-----GCTCCTCGAGCGCCTATACAAGTACTCCTTGT-CTCTCGGCATGGAGC-
   1460
            1470
                          1480
                                    1490
                                              1500
                                                         1510
                                                                   1520
                 1270
                           1280
                                      1290
                                                 1300
     1260
                                                           1310
   CCAAAGCCGGTGGCC--ACCCCTGTCGATA-@TCTTGAGGGA-CGGTAGCGACGACCGTGCTTTT----CGTG
             CCCTAGTCGAGGTCCAGAACACCGAGGAGATGGCCACAGCCATCAAGCTCGGCG~CCAAGGTTATCGGCGTC
      1530
                1540
                          1550
                                    1560
                                              1570
                                                         1580
 1320
          X
   AACTGCAG
    111 1 1
   AACAACCGCAATCTCGAG
     1600 X
               1610
7. LOW344-FIG1. SEQ
  BLYAMYAA
              Barley alpha-amylase type A isozyme mRNA, complete
LOCUS
                         1588 bp ss-mRNA
                                                    PLN
            BLYAMYAA
            Barley alpha-amylase type A isozyme mRNA, complete cds.
DEFINITION
ACCESSION
            J01236
KEYWORDS
            alpha-amylase; amylase.
SOURCE
            Barley (Hordeum vulgare L. cv. Himalaya 1979 crop) aleurone cell
            stimulated with gibberellic acid, cDNA to mRNA, clone E.
  ORGANISM
            Hordeum vulgare
            Eukaryota; Plantae; Embryobionta; Magnoliophyta; Liliopsida;
            Commelinidae; Cyperales; Poaceae; Hordeum; vulgare.
REFERENCE
            1 (bases 1 to 1582)
  AUTHORS
            Rogers, J. C. and Milliman, C.
  TITLE
            Isolation and sequence analysis of a barley alpha-amylase cDNA
            clone
   JOURNAL
            J. Biol. Chem. 258, 8169-8174 (1983)
  STANDARD
            full staff_review
COMMENT
            [1] suggests that there are two alpha-amylase genes in barley
            aleurone cells (types A and B), and that expression of these genes
            is affected differently by gibberellic acid. It is likely that
            alpha-amylase contains a signal peptide and a mature peptide.
            latter would start at about position 172 (by comparison with
            alpha-amylase type B). A poly-A signal is located at positions
            1557-1562.
FEATURES
                     Location/Qualifiers
     mRNA
                     1, . 1588
                     /note="a-amyl mRNA"
     CDS
                     97. . 1413
                     /note="alpha-amylase type A, EC 3.2.1.1"
BASE COUNT
                344 a
                                  480 g
                                           280 t
                         484 C
ORIGIN
            95 bp upstream of Ncol site; chromosome 1.
Initial Score
                ==
                     107
                          Optimized Score =
                                                592
                                                     Significance = 10.89
Residue Identity =
                                                     Mismatches
                     50%
                         Matches
                                           ===
                                                718
                                                                      533
                     171
Gaps
                          Conservative Substitutions
                                                                        0
                                                    40
                     10
                                20
                                          30
                                                                 50
             CTGCAGCCTGACTCGGC-ACCAGTCGCTGCAAGCAGAGTCGT--AAGCA-ATCGCAAGGGG
                  1 1 1 1 1
                            ;
                                 1 1 1 1 1 1 1
                                                         GGTTGGCGTCCGGCCACC-AAGTCCTCTTTCAGGGGTTCAACTGGGAGTCGTGGAAGCAGAGCGGCGGGTGG
      160
                170
                                               200
                                                                   220
             X
                           130
                                     190
                                                         210
               70
                         80
                                   90
                                            100
                                                       110
     -CAGCATGCAAACGAGAAGGGTT&TGC11CAA&TCTGCGGCCGCAGGAACTC-TGCTGGGCGGCCTGGCTGG
```

11 111

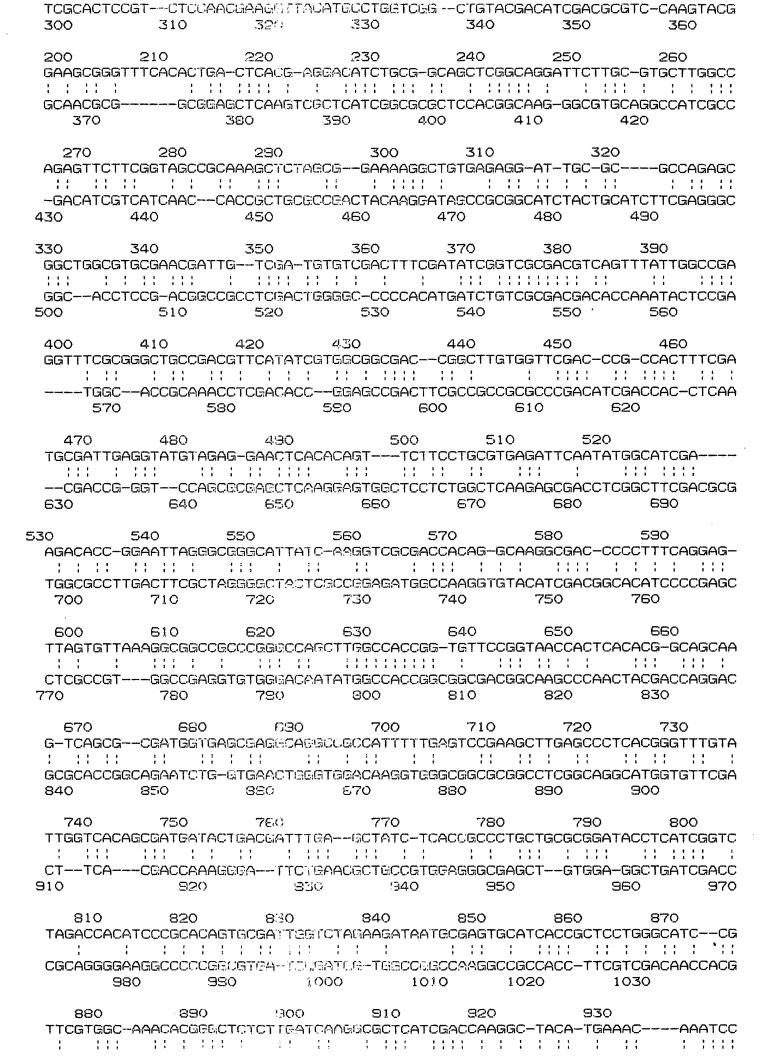


CCGCAGGGGAAGECCCCCGGCGTGA-TGGGATGG-TGGCCGGCCAAGGCCGCCACC-TTCGTCGACAACCAC GTTCGTGGC-AAACACGGGCTCTCTTGATCAAGGCGCTCATCGACCAAGGC-TACA-TGAAAC----AAATC GATACAGGCTCCACGCAGGC-CATGTGGCCATTCCCCTC--CGACAAGGTCATGCAGGGCTACGCGTACATC CTCGTTTCGAATGACTGGCTGTTCGGGGTT-TTCGAGCTATGTCACCAACATCATGGACG-TGATGGATCGCG CTCACCCACCCGGCATCCCATGCATCTTCTACGA-CCATTTCTTCAAC----TGGGGGTTTAAGGA---CC TGAACCCCGACGGGATGGCCTTCATTCCACTCAGAGTGATCCCATTCTACGAG----AGAAGGGCGTCCCA AGATCGCGGCGCTGGTGGCGATCAGGAAGCGCAACGGCATCAC-GGCGACGAGCGCTCTGAAGATCCTCATG CAGGAA-ACGCTGCCAGGCATCACTGTSACTAACCCGGCGGGTTCTGTGTCACCGACTTGCCGTGCATG-A CACGAAGGAGATGCC-TACGTCGCCGAGA-TA-GACGGCAAGGTGGTG-GTGA-AGA-TCG-GGTCCAGGTA CGCCATCTGGATCCT--TCCACGCAGCCGCCACTATTCCCCG--TCAAGATACCGAACGATGAAGTC-GCGC CGACGTCGGGGCGGTCATCCCGGC--CGGGTTCGTGACCTCGGCACGCGCAACG-ACTACGCCGTCTGGGA ATCGATCGAT----AGGCA--TCTTCAA-TGTGATCAGGGCTGC-CACCTCC-AAAGCCGGT-GGCCACCCC TGTC--GATAGT-CTTGAGGGAC----GGTAGCGACGACCG---TGCTTTTC-GTGAACTGCAG TTTCCTGA-AGTACATGATTCACTTCTGGTATTCACG-CGGATATGATTAACTATGTATACCTGTACCCAAA AT 8. LOW344-FIG1. SEQ **PDUMER** Plasmid pDU1358 (from S. marcescens) mercurial resi LOCUS 2153 bp ds-DNA BCT 15-MAR-1988 DEFINITION Plasmid pDU1358 (from S. marcescens) mercurial resistance (mer) operon encoding organomercurial lyase (merB), mercury resistance protein (merD), complete cds, and mercury reductase (merA), 3' end. ACCESSION M15049 KEYWORDS antibiotic resistance; mercuric reductase; mercury resistance; organomercurial lyase; transport protein. Plasmid pDU1358 (multi-antibiotic resistance IncC incompatibility) SOURCE DNA (from Serratia marcescens), clone pHG6. ORGANISM Plasmid pDU1358 Prokaryota; Bacteria; Plasmid pDU1358. 1 (bases 1 to 2153) REFERENCE **AUTHORS** Griffin, H. G., Foster, T. J., Silver, S. and Misra, T. K. TITLE Cloning and DNA sequence of the marcuric and organomercurial

```
resistance determinants of plasmid pDU1358
  JOURNAL
           Proc. Natl. Acad. Sci. U.S.A. 84, 3112-3116 (1987)
  STANDARD
           full staff review
COMMENT
           Computer-readable sequence for [1] kindly provided by S. Silver,
FEATURES
                  Location/Gualifiers
                  < 1..359</p>
     CDS
                  /note="mercuric reductase (merA; AA at 3)"
     CDS
                  /note="organomercurial lyase (merB)"
                  1124..1489
     CDS
                  /note="mercury resistance protein (merD"
BASE COUNT
              391 a
                     641 c
                             680 g
                                     441 t
ORIGIN
          Unreported.
Initial Score
              272
                  95 Optimized Score =
                                         580 Significance = 10.17
Residue Identity =
                  50% Matches
                                         727 Mismatches
                                                            511
Gaps
                  194 Conservative Substitutions
                                                             0
                  10
                           20
                                    30
                                                40
                                                        50
           CTGCAGCCTGACTC-GGCACCAGTCGCTGCAAG-CAG--AGTC-GTAAGCAATCGCA--AGG
                  TGGCGGTGTCCTTGGTATTGCCGCAGGAAGCAGCCGACGTTCGTCAGTCCTTCTGTTGCCATGTACATTTCT
     800
           X 810
                           830
                  . 820
                                       840
                                                850
     60
               70
                         80
                                 90
                                         100
                                                 110
                                                          120
   GGGCAGC-ATGCAAACG--AGAAGGGTTGTGCTCAAGTCTGCGGCCGCAGGAACTCTGCTGGGCGGCCTGGC
     1 1 1 1
                                                  1 1 1 1 1 1
   TTGCATCTGTCCCGACGGCGGAAGACTGGGCCTCCAAGCATCAAGGATTGGAA---GGATTGGCGATC-GTC
    870
            880
                    890
                             900
                                     910
                                                 920
                       150
                               160
                                        170
                                                 180
   TGGGTGCGCGACGTGGCTGGATC-GATCGGCACAGGCGATCGGATCAATACGTG-CGCGTCCTATCACAATC
    AGTGT---CCACGAGGCT--TTCGGCTTGGGCCAGGAG-TTTAATCGACATCTGTTGCAGACCATGTC-ATC
        940
                  950
                          960
                                    970
                                            980
                                                     990
      200
                210
                        220
                                230
                                         240
                                                 250
                                                             260
   TCTGA-AGCG-GGTTTCACACTGACTCACGAGGACATCTGCGGCAGCTCGGCAGGATTC-TTGCGTGC---T
     TAGGACACCGTGATCGGATATCGACCCA--ATG--TTCTACGGCACCGGCATCGGATTCGCAGCGCGCGAT
1000
         1010
                 1020
                           1030
                                     1040
                                              1050
                                                      1060
         270
                 280
                           290
                                     300
                                                310
   TGGCCAGAGTTCTTCGGTAGCCGCAAAGC--TCTAGCG--GAAAAG--GCTGTGAGAGGATTGCGCGCCAGA
              TGAACTCGGCGAAACGGTATATGCATTGCCGTGAACCGACCAAAAGGAGGTGTTCGATGAACGC-CTACACG
  1070
          1080
                   1090
                          1100
                                   1110
                                            1120
                                                     1130
             340
                     350
                              360
                                      370
                                               380
   GCG-GCTGGCGAACGATTGTCGATGTCGACTTTCGATATCGGTCGCGACGTCAGTTTATTGGCCGA
   GTGTCCCGGCTGGCCCTTGA-TGCCG---GGGTGAGCGTGC-ATATCGTGCGCGACTAC---CTGCTGCGCG-
 1140
                   1150
                             1170
                                      1180
                                               1190
   400
            410
                     420
                             430
                                     440
                                              450
                                                       460
   GGTTTC-GCGGGCTGCCGACGTTCATATCGTGGCGGCGACCGGCTTGTGGTTCGAC-CCGCCACTT---TCG
   GATTGCTGCGGCCAGTCG-CCTGCACCACG-GGTGGCTA-CGGCCTGTTCGATGACGCCGC--CTTGCAGCG
      1210
               1220
                        1230
                                  1240
                                          1250
                                                   1260
                 480
                            490
                                     500
                                             510
   A-TGCGAT---TGAGGTATG---TAGAGGAACTCACACAGTTCTTCCTGCGTGAGATTCAATATGGCATCGA
           1 1 1
   ACTGTGCTTCGTGCGGCCGCCTTCGAGG---CGGGCATCGGCCT--CGGCG----CATTGGCGCGGCTGTGCC
1270
                 1290
         1280
                           1300
                                    1310
```

530				560	570	580	590	
A	GACACCGGA	ATTAGGGCGC	BIGC-ATTATO:	PAGGTOGO	GACCA-CAG	GCAAGGCG	ACCCCCTTT	C-AGGAG
	1 1 1 111	: ::::	: ::		: ::::	1: 1:	: : ::	1 11 1
G	GGCGCTGGA-	TGCGGCG/	PACTGCGATG	PAACTGCC	GCGCAGCTT	GCTGTGCT	GCGTCAGTT	CGTCGAACG
	1340	1350	1360	137	0 13	80	1390	1400
					-			
	600	610	620	6	30	540	650	660
			3CCC613G					
1	1 1 11	1 1111	1 11			11 111	1 1 1 1 1 1	
~							111111	111 11
U			ATCTGGAAG					
	1410	1420	1430	1440	14	50	1460)
	670	_		690	700	710		720
			GTGAGCG					GAGC
	:: ::	11 111			111 1	:: :	1:::: ::	; ;
A)	TGCGGAGAG	TTTGCCATG?	AACAGCCCCG	AGCGCATG(CCGGCC	GAGACACA	CAAGCCGTT	CACCGGCTA
1470	1480) 149	30 150	00	1510	1520	153	SO
	730	74	4O '	750	760	770	· 780	790
C	CTCACGGGT	TTG-TATTG	GT-CACAGCG	ATGATACTI	GACGATTTG	AGCTATCT	CACCGCCCT	GCTGCGCGG
	1 1 1 1 1 1					111		
			STGCTCACC-					
154			560	1570	1580	159		600
1	0 15.	50 10	JGQ	1570	1360	100	, ,	300
	000	01/	2	_	222	040	050	000
0	800	810			830 ATTO STA	840	850	860
H			CACATCCCGC					GCATCACCG
		1 1 1 1			1 1 1 1 1	1 1 1	: ::	
			ATCGGACAGC				GA-CCG	-GCTTGTTTG
16	10 16	520	1630	1640	1650	1660		1670
	870	88	80 8:	B0	900	(310	920
C	TCCTGGGCAT	TCCGT-TCG1	TGGCAAACAC	BESCTOTO:	TTGA	TCAAG-GC	GCTCATCGA	CCAAGGCTA
	1111	:: :: ::	111	: : : : : : : : : : : : : : : : : : : :	; ;;	::: : ::	::: :	1111
	TCCTG	TCTGTGACG	CGGCTGCTGC	GGCCTTC/	AGAGGTCGA	TCATGAGC	GCTTCCCAG	CCAATTG
	1680	1690	1700	5 1'	710	1720	1730	1740
				-				
	930	940	950	960	o :	970	980	990
C			TTTCGAATGA		_			
Ο,	111 111							: :: :
^	VIC-CVCVC.		raacaca-aa					
1-1	1750				1780			
	1 7 30	J 176	50 1	770	1760	1730	1800	1810
	1000	1010		~	1070	1040	1050	
_			1020					
اها	GACGIGAIG		BAACCCCGAC					
	11 11 1		;;;;					
CI			BATTGTCGGC					
	1820	0 183	3O 19	4O	1850	1860	187	' O
	1070	0 108	30	1090	110		1110	1120
A			30 CAAACGC-TG			0		
		GTCCCACAGO	CAAACGC-TG	CCAGGC		O G-TGAC-TA	ACCCGGCG	icGGTTC
	GAAG——GGCO ¦ ¦ ¦	STCCCACAGO	CAAACGC-TG	CCAGGC(ATCACT	O G-TGAC-TA	ACCCGGCG	CGGTTC
T	GAAGGGC0 ¦ ¦ ; GTTGCCGCC0	GTCCCACAGO 	CAAACGC-TG- 	CCAGGC/ ; ; ; >TCGAATC	ATCACTI	O G-TGAC-TA ! !! ! GATGCCGGA	ACCCGGCG 	CGGTTC CATGGGTGC
T	GAAG——GGCO ¦ ¦ ¦	GTCCCACAGO 	CAAACGC-TG- 	CCAGGC/ ; ; ; >TCGAATC	ATCACT	O G-TGAC-TA ! !! ! GATGCCGGA	ACCCGGCG	CGGTTC
T	GAAGGGCC ; ; ; ; GTTGCCGCCC O 189	GTCCCACAGO GGCCAGTTC O 15	CAAACGC-TG- ; ; ;; CCTGCCCGTG BOO 19	CCAGGC/ ;;;; ATCGAATC(310	ATCACTI :: :: GTCCGGCCTI 1920	O G-TGAC-TA ! !! ! GATGCCGGA 1930	AACCCGGCG 	CGGTTC CATGGGTGC
Т(188¢	GAAGGGC0 ; ; ; ; GTTGCCGCC0 O 189	GTCCCACAGO	CAAACGC-TG- : : :: CCTGCCCGTG 300 19	CCAGGCA ; ; ; ATCGAATCG B10	ATCACTI 	0 G-TGAC-TA :	AACCCGGCG 	CGGTTC : CATGGGTGC 1950
T(1880 T(GAAGGGCO	STCCCACAGO	CAAACGC-TG- ; ; ;; CCTGCCCGTG BOO 19 D 1150 BCATGACGCC	CCAGGCA ; ; ; ATCGAATCA 310 1 ATCTGGATA	ATCACTI	O G-TGAC-TA ; ;; ; GATGCCGGA 1930 1170 CAGCGGCCA	AACCCGGCG 	CGGTTC : ::: : CATGGGTGC 1950
T(1880 T(;	GAAGGGCO GTTGCCGCCO 0 189 1130 GTGTCACCGA	GTCCCACAGO	CAAACGC-TG- : ; ;; CCTGCCCGTG 900 19 0 1150 CATGACGCC ; ;;;	CCAGGC/ ;;;; ATCGAATC B10 D 1 ATCTGGAT(;;;;	ATCACTI :: :: GTCCGGCCTI 1920 160 CCTTCCACGI : : :	O G-TGAC-TA ; ;; ; GATGCCGGA 1930 1170 CAGCGGCCA	AACCCGGCG 	CGGTTC : ::::: CATGGGTGC 1950CCGT
T(1880 T(;	GAAGGGCC	GTCCCACAGO	CAAACGC-TG- :	CCAGGCA ; ; ; ATCGAATCA 310 1 ATCTGGATA ; ;;;; A-CTGG-CA	ATCACTI	O G-TGAC-TA I II I GATGCCGGA 1930 1170 CAGCGGCCA II III CA-TGGCAA	AACCGGGGG 	CGGTTC
T(1880 T(;	GAAGGGCO GTTGCCGCCO 0 189 1130 GTGTCACCGA	GTCCCACAGO	CAAACGC-TG- :	CCAGGC/ ;;;; ATCGAATC B10 D 1 ATCTGGAT(;;;;	ATCACTI	O G-TGAC-TA ; ;; ; GATGCCGGA 1930 1170 CAGCGGCCA	AACCGGGGG 	CGGTTC : ::::: CATGGGTGC 1950CCGT
T(1886 T(: T(GAAGGGCC	STCCCACAGO	CAAACGC-TG- ; ; ;; CCTGCCCGTG/ BOO 19 CATGACGCC/ ;;; CAAATGCGCG/ P7O	CCAGGCA ; ; ; ATCGAATCO 310 1 ATCTGGATO ; ;;; A-CTGG-CO	ATCACTI II	O G-TGAC-TA I I I GATGCCGGA 1930 1170 CAGCGGCCA II III CA-TGGCAA	AACCCGGCG	GCGGTTC : :::::: GCATGGGTGC 1950CCGT ::::: GCTGGCCGT
T(1886 T(; T(GAAGGGCC	GTCCCACAGO	CAAACGC-TG- : ; ;; CCTGCCCGTG: 800 19 0 1150 CCATGACGCCG: 1 ;;; CAAATGCGCGG	CCAGGCA ;;;; ATCGAATCO 310 1 ATCTGGATO ;;;;; A-CTGG-CO 1980	ATCACTI II II II ATCOGGCCTI ASSO ASSO ASSO ASSO ASSO ASSO ASSO ASS	0 G-TGAC-TA GATGCCGGA 1930 1170 CAGCGGCCA CA-TGGCAA 2000	AACCCGGCG	CGGTTC : :::::: CATGGGTGC 1950CCGT ::::: GCTGGCCGT
T(1886	GAAGGGC0	GTCCCACAGO	CAAACGC-TG-	CCAGGCA ;;; ATCGAATCG B10 1 ATCTGGATG CTGG-CG LS20 CSATCGATG	ATCACTI II	0 G-TGAC-TA GATGCCGGA 1930 1170 CAGCGGCCA CA-TGGCAA 2000 1240 CAATGTGA	AACCCGGCG	CGGTTC : ::::: CATGGGTGC 1950CCGT ::::: GCTGGCCGT 250 GCCACCTCC
T(1886) T(1876) 1 C(1876)	GAAGGGC0	GTCCCACAGO GGGCCAGTTO 1140 ACTTGCCGTO -CCTGCCGTO SO 19 ACGA-TGAA	CAAACGC-TG- :	CCAGGCA ; ; ; ATCGAATCA B10 D 1 ATCTGGATA ; ; ; ; A-CTGG-CA 1980 1220 CGATCGATA ; ; ; ; ;	ATCACTI II II II ATCCGGCCTI ATCCGGCCTI ATCCGGCCTI ATCCGGCCTI ATCCGGCTGCTGGI ATCCGGGTGCTGGI ATCCGGGTGCTGGI ATCCGGGTGCTGGI ATCCTTI	DOG TGAC-TA 	AACCCGGCG I I I I I I I I I I I I I I I I I I I	GCGGTTC
T(1886) T(1876) 1 C(1876)	GAAGGGCC	GTCCCACAGO	CAAACGC-TG-	CCAGGCA ; ; ; ATCGAATCA B10 D 1 ATCTGGATA ; ; ; ; A-CTGG-CA 1980 1220 CGATCGATA ; ; ; ; ;	ATCACTI II	DOG TGAC-TA 	AACCCGGCG I I I I I I I I I I I I I I I I I I I	GCGGTTC

```
1270
                         1280
                                    1290
                                               1300
                                                         1310
                                                                   1320
    AAAGCC-GGTGCCCACCCCTGTCGATAB--TCTTGAG-GGACGGTAGCGACGACCGTGCTTTTCGTGAACTGC
     1 1 1 1 1
    GATGCCGGGTTGCCCGCGCGCGTATCTTWAAATTGAGCTGACCGAATCG----GTTGCGTTCGGTGATCCGG
            2090
  2080
                      2100
                                2110
                                          2120
                                                         2130
                                                                   2140
     X
    AG
    CGATCTTC
     2150
9. LOW344-FIG1. SEQ
   BLYAMY2
               Barley (H. vulgare) alpha-amylase 2 mRNA, complete
 LOCUS
                          1588 bp ss-mRNA
                                                     PLN
             Barley (H. vulgare) alpha-amylase 2 mRNA, complete cds.
 DEFINITION
 ACCESSION
             M17128
 KEYWORDS
             alpha-amylase.
 SOURCE
             Barley (H. vulgare cv Sundance), cDNA to mRNA, clone E.
   ORGANISM
             Hordeum vulgare
             Eukaryota; Plantae; Embryobionta; Magnoliophyta; Liliopsida;
             Commelinidae; Cyperales; Poaceae; Hordeum; vulgare.
 REFERENCE
             1 (bases 1 to 1588)
   AUTHORS
             Knox, C. A. P., Sonthayanon, B., Chandra, G. R. and Muthukrishnan, S.
   TITLE
             Structure and organization of two divergent alpha-amylase genes
             from barley
   JOURNAL
             Plant Mol. Biol. 9, 3-17 (1987)
   STANDARD
             full staff entry
             Draft entry and computer-readable sequence for [1] kindly provided
 COMMENT
             by S. Muthukrishnan, 22-SEP-1987.
FEATURES
                     Location/Qualifiers
     mRNA
                      1..1588
                      /note="alpha-amylase 2 mRNA (alt.)"
     mRNA
                     2. 1588
                     /note="alpha-amylase 2 mRNA (alt.)"
     mRNA
                     3..1588
                     /note="alpha-amylase 2 mRNA (alt.)"
     mRNA
                     4. . 1588
                     /note="alpha-amylase 2 mRNA (alt.)"
     CDS
                     97..1413
                     /note="alpha-emylase 2"
 BASE COUNT
                343 a
                         484 C
                                  480 g
                                           281 t
ORIGIN
             380 bp upstream of SStI site.
Initial Score
                     107
                          Optimized Score =
                                                580
                                                     Significance = 10.17
Residue Identity =
                     50%
                          Matches
                                                714
                                                     Mismatches
                                                                      538
Gaps
                     169
                          Conservative Substitutions
                                                                         0
                      10
                                20
                                          30
                                                    40
                                                                 50
             CTGCAGCCTGACTCGGC-ACCAGTCGCTGCAAGCAGAGTCGT--AAGCA-ATCGCAAGGGGG
                    11 1 1 1 1
                                 111
                                      ; ; ;
                                                1 1 1 1 1 1 1
                                                         GGTTGGCGTCCGGCCACC-AAGTCCTCTTTCAGGGGTTCAACTGGGAGTCGTGGAAGCAGAGCGGCGGGTGG
       160
             X
                170
                            130
                                      150
                                               200
                                                         210
                                                                   220
               70
     60
                         80
                                   90
                                            100
                                                       110
    --CAGCATGCAAACGAGAAGGGTTETGCTCAAGTCTGCGGCGCAGGAACTC-TGCTGGGCGGCCTGGCTGG
               : :
                                   111 111
   TACAACATGATGATGGCCAAGGTCGACGACATCGCCGCCGTCGGAGTCACCCACGTCTGGCTGCCACCGCCG
     230
               240
                         250
                                   260
                                             270
                                                       280
                                                                 290
   130
             140
                       150
                                 160
                                             170
                                                       180
                                                                 190
   GTGCGCGACGTGGCTGGATCGGCACAGGC--GATCGGATCAATACGTGCGCGTCCTATCACAATCTCT
```



```
1040
                   1000
                            1070
                                       1080
                                                1090
 940
          950
                   260
                             970
                                      980
                                               990
   TOGTTTCGAATGACTGGCTGTTCEGGTT-TTCGAGCTATGTCACCAACATCATGGACG-TGATGGATCGCGT
                          1 1 1
                                                 1 1 1
   TCACCCACCCGGCATCCCATGCATCTTCTACGA-CCATTTCTTCAAC----TGGGGGTTTAAGGA---CCA
 1110
          1120
                   1130
                            1140
                                      1150
                                                   1160
1010
         1020
                   1030
                           1040
                                    1050
                                             1.060
                                                           1070
   GAACCCCGACGGGATGGCCTTCATTCCACTGAGAGTGATCCCATTCTACGAG-----AGAAGGGCGTCCCAC
             1 1 1 1 1 1 1 1
                            : :
                                 1 1 1 1
                                           1 1 1 1 1 1
                                                       1 1 1 1
   GATCGCGGCGCTGGTGGCGATCAGGAAGCGCAACGGCATCAC-GGCGACGAGCGCTCTGAAGATCCTCATGC
       1180
                1190
                         1200
                                  1210
                                            1220
                                                     1230
   1080
             1090
                      1100
                               1110
                                        1120
                                                 1130
   AGGAA-ACGCTGCCAGGCATCACTGTGACTAACCCGGCGCGCGTTCTGTGTCACCGACTTGCCGTGCATG-AC
                ACGAAGGAGATGCC-TACGTCGCCAAGA-TA-GACGGCAAGGTGGTG-GTGA-AGA-TCG-GGTCCAGGTAC
      1250
                1260
                         1270
                                    1280
                                             1290
                                                         1300
   1150
            1160
                       1170
                                1180
                                           1190
                                                    1200
                                                              1210
   GCCATCTGGATCCT---TCCACGCAGCGCCACTATTCCCCG---TCAAGATACCGAACGATGAAGTC-GCGCA
              1 111 11
                         1 1 1
                                   GACGTCGGGGCGGTGATCCCGGC---CGGGTTCGTGACCTCGGCACACGCAACG-ACTACGCCGTCTGGGAG
                   1330
                              1340
                                       1350
                                                1360
                                                          1370
      1220
                    1230
                              1240
                                        1250
                                                  1260
   TCGATCGAT----AGGCA--TCTTCAA-TGTGATCAGGGCTGC-CACCTCC-AAAGCCGGT-GGCCACCCCT
                     : : :
                                                  1380
           1390
                    1400
                             1410
                                      1420
                                               1430
                                                        1440
                1290
      1280
                            1300
                                        1310
                                                  1320
   GTC--GATAGT-CTTGAGGGAC----GGTAGCGACGACCG---TGCTTTTC-GTGAACTGCAG
                          : 1 1 1
       ; ;
; ;
                                TTCCTGA-AGTACATGATTCACTTCTGGTATTCACG-CGGATATGATTAACTATGTATACCTGTACCCAAAA
1450
          1460
                   1470
                                      1490
                                               1500
                            1480
                                                        1510
   T
1520
10. LOW344-FIG1. SEQ
   BPECYADE
              Bordetella pertussis cyaD gene 3'region and cyaE g
LOCUS
           BPECYADE
                       2040 bb ds-DNA
                                               BCT
DEFINITION
           Bordetella pertussis cyaD gene 3'region and cyaE gene; proteins
           necessary for transport of calmodulin-sensitive adenylate cyclase-
           haemolysin (cyclolysin).
ACCESSION
KEYWORDS
           adenylate cyclase; cya gene; cyaD gene; cyaE gene; hemolysin;
           secreted protein; toxin.
SOURCE
           Bordetella pertussis.
  ORGANISM
           Bordetella pertussis
           Prokaryota; Bacteria; Gracilicutes; Scotobacteria; Aerobic rods and
           cocci; Alcaligenaceae; Bordetella; pertussis.
REFERENCE
           1 (bases 1 to 2040)
  AUTHORS
           Glaser, P., Sakamoto, H., Bellalou, J., Ullmann, A. and Danchin, A.
  TITLE
           Secretion of cyclolysin; the calmodulin-sensitive adenylate cyclase
           - haemolysin bifunctional protein of Bordetella pertussis
  JOURNAL
           EMBO J. 7, 3997-4004 (1988)
  STANDARD
           simple automatic
COMMENT
           *source: strain 43323.
                                 see Y00545 for upstream cya gene; cya
           operon is organized cyaABDE, cyaB (712 aa) is initiated 78 bp
```

downstream of cyaA stop; cyaB stop overlaps with cyaD initiation

```
(440 aa).
        EMBL features not translated to GenBank features:
               from
                    to description
                           put. rRNA-binding site
        RBS
                344
                    347
              Location/Qualifiers
FEATURES
   CDS
              /note="cyaD polypeptide (AA at 3)"
   CDS
              /note="cyaE polypeptide (AA 1-474)"
BASE COUNT
          282 a
                679 c
                     770 g
                           309 t
ORIGIN
Initial Score =
              82 Optimized Score =
                               590 Significance = 10.17
Residue Identity =
              51% Matches
                               732 Mismatches
                            :::::
              197 Conservative Substitutions
Gaps
                                               0
               20
                       . 30
                               40
                                     50
   CTGCAGCCTG-ACTCGGCACCAGTC---GCT-GCAAGCAGAGTCGTAAGCAATCGCAAGG-GGGC-AGCAT
   20
                     30
                           40
     70
           80
                  90
                       100
                              110
                                    120
                                           130
  GCTACCGTCAA-GGTCG-GC----GCCTACGACTATACGAAGTACGGAACGCTCGAAGGCAAGGTGTTGTAT
       80
                 90
                       100
                             110
           150
                  160
                        170
                              180
                                      190
  GTGGCT--GGAT-CGATCGCACAGGCGATCGGATCAATACGTGCGCGT--CCTATCACAATCTCTGAAGCG
  GTGTCTCCGGATACGGTGGTC---GACGACCG--CCAACA-GCACTCGTACCGCGTGACGATCGC----GCT
  140
         150
                 160
                        170
                                180
                                      190
            220
                     230
                            240
                                    250
     210
  GGTTTCACACTGAC--TCACGAGGAC----ATCTGCGGCAGCTC--GGCAGGATTCTTGCGTGCTTGGCCAG
  200
        210
              220
                     230
                           240
                                  250
          280
                 290
                        300
                               310
                                     320
  AGTTCTTCGG--TAGC-CGCAAAGCTCTAGCG-GAAAAGGCTGTGAGAGGATTGCGCGCCAGAGCGGC--TG
  A--TATCCGGACCGGCTCGCGGCGCCTCATCGAGTATCTGCTCAGCCCGG-TGGCGCGGCATGCCGGCGAAA
   270
          280
                290
                       300
                             310
                                    320
                                           330
      340
            350
                   360
                         370
                                380
                                      390
                                             400
  GC-GTGCGAACGATTGTCGATGTCGACTTTCGATATCGGTCGCGACGTCAGTTTATTGGCCGAGGTTTC
  340
         350
                         370
                               380
                  350
                                      390
     410
           420
                  430
                           440
                                  450
                                        460
  GCGGGCTGCCGACGTTCATATCGT-GGCGGCG----ACCGGCTTGTGGTTCGACCCGCCACTTTCG-ATGCG
         410
          420
                 430
                       440
                              450
                                      460
 470
                 490
                         500
         480
                                510
                                      520
  ATTGAGG--TATGTAGAG---GAACTCACACAG-TTCTTC-CTGCGTGAGATTCAATATGGCATCGAAGAC
        TTCGAGGGCCAGGCGCCTGCCCTCTCGTGGCCTTGTCCGCCGCCGGCGGATC----GGC-TCGACGAC
          490
                       510
                              520
                                        530
                                               540
    480
                 500
     540
            550
                  560
                           570
                                  580
                                        590
                                               600
```

ACCGGAATTAGGGCG-GGCATTATCAAGGTCGCG-----ACCACAGGCAAGGCGACCCCCTTTCAGGAGTTAG